A HIERARCHICAL APPROACH FOR HUMAN AGE ESTIMATION

Pavleen Thukral[1], *Kaushik Mitra*[2] *and Rama Chellappa*[3]

[1]Poolesville High School, MD
[2]ECE, Rice University, TX
[3]ECE and UMIACS, University of Maryland, College Park, MD

ABSTRACT

We consider the problem of automatic age estimation from face images. Age estimation is usually formulated as a regression problem relating the facial features and the age variable, and a single regression model is learnt for all ages. We propose a hierarchical approach, where we first divide the face images into various age groups and then learn a separate regression model for each group. Given a test image, we first classify the image into one of the age groups and then use the regression model for that particular group. To improve our classification result, we use many different classifiers and fuse them using the majority rule. Experiments show that our approach outperforms many state of the art regression methods for age estimation.

***Index Terms*—** Age Estimation, Combined Regression and Classification

1. INTRODUCTION

Simulation and recognition of faces as one ages has recently been gaining popularity as a topic in biometrics literature [1]. Understanding facial aging is very important not only as a scientific problem, but also as a means to solving a sub problem in facial recognition, where the claimed identity and enrolled face may exhibit various differences on the feature space due to a large difference in ages of the subjects. Analyzing how structure, when using geometric analysis, and how appearance, when using textural analysis, vary over aging yields important information for designing face recognition systems.

This paper considers the problem of automatic human age estimation from images. Age estimation is generally formulated as a regression problem between features extracted from the face image and the age variable [2, 3, 4, 5]. It has been shown that facial geometry, characterized by 2-dimensional landmarks extracted from the face images (see Fig.1) is a strong indicator of age progression [2]. However, the age progression across different age groups is quite different. For example, the facial geometry at young ages changes much faster than at old ages. Hence, it is not effective to solve the age estimation problem using a single regression model. We propose a hierarchical approach in which we learn a separate regression model for each age group. Given a test image, we first classify the image in the appropriate age group and then use the regression model for that age group. We use the geometric features proposed in [2] as our feature vector and the Relevance Vector Machine (RVM) regression [6] as the regression technique. For classifying the test image in the proper age group, we use many different classifiers such as μ-SVC [7], Partial least squares (PLS) [8], Fisher Linear Discriminant, Nearest Neighbor, and Naive Bayes [9] and fuse them using the majority rule. Our experiments on the FG-Net dataset [10] shows that the hierarchical approach provides much better age estimation results than using a single regression model for all ages.

Fig. 1. 2D facial landmarks on an image from the FG-Net dataset.

The paper is organized as follows: in section 2, we propose our hierarchical approach for age estimation. In section 3, we present our experimental results on the FG-Net dataset and section 4 concludes the paper.

2. HIERARCHICAL AGE ESTIMATION

There are three major steps in our approach: feature extraction, learning regression model for each of the age groups and classifying the test image into the various age groups. We discuss these steps in the following sub-sections.

2.1. Feature Extraction

We extract geometric features from the face image as proposed in [2]. These features are based on landmark 2D points such as corners/extremities of eyes, mouth, nose, etc. which can be reliably located on most faces. There exist several automatic methods to locate facial landmarks which work well on constrained images such as passport photos [11]. However, these landmark points are very sensitive to affine transformation arising due to small changes in view. To take care of this problem, the collection of landmark points in an image is treated as a point in the Grassmanian manifold. First the manifold mean of all such points are computed (the mean represents the collection of landmark points of an average face) and then these points are projected onto the tangent plane at the mean. Any given face is then parameterized by the velocity vector (on the tangent plane) that transforms the average face to the given face in unit-time. These velocity vectors are the features that are used for regression. For more details see [2].

2.2. A Separate RVM Regression for Each Age Group

The goal of regression is to learn the functional relation between two sets of variables: the independent variable $\mathbf{x}$ and the dependent variable y, using many example pairs $(\mathbf{x_i}, y_i), i = 1, 2, , N$. RVM regression [6] assumes that the y and $\mathbf{x}$ are related as follows:

$$y = \sum_{i=1}^{N} w_i k(\mathbf{x}, \mathbf{x_i}) + w_0 + e, \quad (1)$$

where $k(\mathbf{x}, \mathbf{x_i})$ is a kernel function and e is a Gaussian noise variable. The objective is to estimate the weight vector $\mathbf{w} = [w_0, w_1, , w_N]^T$ using the training dataset.

RVM is a Bayesian regression approach where a sparse prior is assumed for the weight vector $\mathbf{w}$. The prior is specified as a hierarchical prior:

$$p(\mathbf{w}|\alpha) = \prod_{i=0}^{N} \mathcal{N}(w_i|0, \alpha_i^{-1}), \quad (2)$$

that is, each component w_i is a Gaussian random variable with mean 0 and variance α_i^{-1}, where the α_i variables, known as the hyper-parameters, are assumed to be uniformly distributed. The true nature of this hierarchical prior becomes apparent once we integrate over the hyper-parameters:

$$p(w_i) = 1/|w_i|, \quad (3)$$

which is indeed a sparse prior. Given a training set of observations $(\mathbf{x_i}, y_i)$, the weight vector $\mathbf{w}$ is solved using the maximum a posteriori (MAP) criterion. And once we obtain $\mathbf{w}$, we can use it to predict the value of y for any new $\mathbf{x}$ using the RVM model (1). For more details about RVM regression see [6].

We learn a separate RVM regression model for each age group. The number and the age-range of the different age groups are decided based on the perceived homogeneity in the age-group and the number of data available for training. Also during the RVM training phase, we allow some of the training samples to overlap between the age groups so that that the models are robust to classification errors that might occur while assigning the test image to the correct age group.

As a choice for the kernel function $k(\mathbf{x_i}, \mathbf{x_j})$ in the RVM model (1), we chose the Gaussian kernel which is of the form:

$$k(\mathbf{x_i}, \mathbf{x_j}) = exp(-\frac{||\mathbf{x_i} - \mathbf{x_j}||^2}{r^2}) \quad (4)$$

The choice of the unknown parameter r is critical in obtaining good regression performance. Hence, during the training phase, we optimize this parameter for the different age groups.

2.3. Classification Into One of the Age Groups

The third part of the hierarchical approach is automatic classification of the test subject into one of the age groups using multiple classifiers. We use five classifiers: μ-SVC [7], Partial Least Squares (PLS) [8], Nearest Neighbor, Naive Bayes and Fisher Linear Discriminant [9]. We use the majority rule to obtain the final classification of our test subject into the proper age group.

3. EXPERIMENTS

We evaluate our proposed approach on the FG-Net dataset [10], which contains 1002 images of 82 subjects at various ages. For this dataset, 68 landmark points are available with each face, see Figure 1. We divide the data into three age groups of $0 - 15, 15 - 30$ and 30+. This division is based on the perceived homogeneity of the geometric growth of human faces with age. For example, there is a rapid geometric growth in the age group $0 - 15$, moderate growth in the age group $15 - 30$ and almost no growth after the age of 30. Similar division into different age groups has been used for age classification in [12, 13]. Another consideration for such a division is the number of images available at the different age groups, see Fig. 2 for the distribution of age in the FG-Net dataset. For appropriately training the RVM regression machines and the classifiers, each age group must have sufficient number of data. In the following paragraphs, we first present our age regression results with ground-truth classification. i.e., when

Age groups	0-15	15-30	30 and above
Optimal r value	0.193	0.178	0.396

Table 1. Optimal scale value r for the Gaussian kernel (4) for different age groups.

there is no error in classifying the test images into the different age groups. And then we present our overall hierarchical age estimation results.

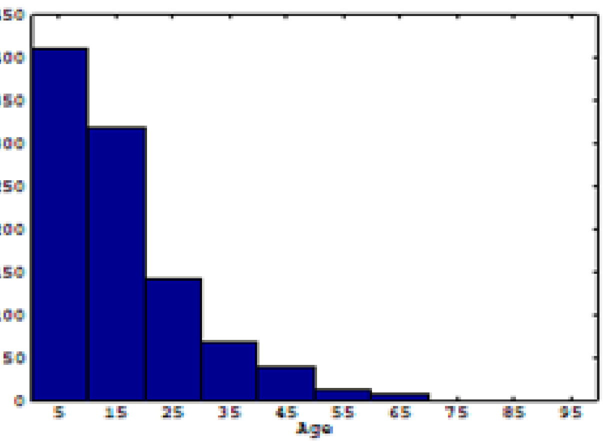

Fig. 2. Distribution of age in the FG-Net dataset.

Age Estimation With Ground-truth Age Group Classification: First we extract the geometric feature vectors (as described in section 2.1) from each of the face images in the dataset. Then for each age group, we learn a separate RVM model. There are two important parameters to learn in the RVM model: the Gaussian kernel width r (4) and the model weight vector $\mathbf{w}$ (1). First we optimize over the parameter r. The r value is an indication of the scale of the distances between the feature vectors. Table 1 shows the optimal r values found for each of the three age groups. With this optimal setting for the r value, we are now in a position to obtain the weight vector $\mathbf{w}$ for each group.

To measure the performance of the proposed approach we perform leave-one-subject-out testing, in which all the images of a person are used as the test set and the rest of the images are used as the training set. We first learn the weight vector $\mathbf{w}$ on the training set and then use it for estimating the age of the test images. We measure the estimation error using the mean absolute error (MAE) metric, defined as MAE$=\frac{\sum_{i=1}^{N}(l_i-\hat{l_i})}{N}$, where l_i is the true age and $\hat{l_i}$ is the estimated age. We obtained an overall (across all age groups) MAE of 3.41 compared to the current state-of-the-art result of 5.07 [5]. Figure 4 shows the predicted age of the test images along with their ground-truth values. It can be seen that the age estimation in the age group 30 and above is worse than the other age groups. This is because our features capture the geometry of the face and there is less variation in the facial geometry in this age group.

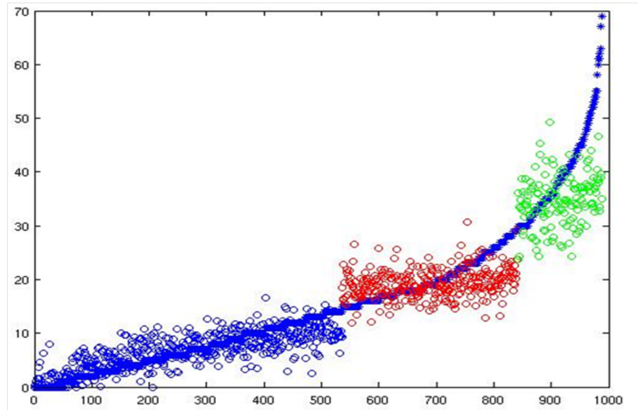

Fig. 3. The estimated age values for the test images are shown as circles and the ground-truth values are shown as asterisks. Here we have assumed that we know the correct age group for each of the test images.

Age estimation using the overall hierarchical framework: In our overall hierarchical framework, we first classify a test image into one of the three age groups by combining the results of the following classifiers in a majority rule: μ-SVC (with $\mu = 0.3$), Fisher Linear Discriminant (FLD), Partial Least Squares (PLS) regression, Nearest Neighbor (NN), and Naive Bayes. We then use the appropriate RVM regression model to estimate the age of the test image. The classification stage of the hierarchical approach is very important. Hence, first we tested the performance of each of the classifiers for their ability to correctly classify the test images. The results are shown in table 2 from which we can conclude that classifier fusion produces better result than any of the individual classifiers. The overall classification rate of 70% is lower than the classification rate obtained in [13], which uses both geometric and textural cues for classification. We then tested the overall approach in the leave-one-subject-out mode. Figure 4 show a plot of the ground truth and predicted age values. Table 3 shows that our approach out-performs many state of the art regression approaches with the exception of RUN1 [14], LARR [5] and warping velocity SVM [2].

4. CONCLUSION AND FUTURE WORK

We propose a hierarchical approach for age estimation, where we divide the face images into different age groups and learn a separate regression for each of the age group. Our experiments show that if the test image can be classified into the correct age group, then the task of age estimation can be performed very accurately. However, with the current classification rate of 70%, we obtain a sub-optimal performance with our overall hierarchical framework. In future, to improve the

Classifier	μ-SVC	PLS	FLD	Naive-Bayes	NN	Majority rule
Accuracy	65.8%	68.3%	55.9%	64.9%	61.6%	70%

Table 2. The percentage of test images classified correctly by the different classifiers. The results show that the classifier fusion produces better result than any of the individual classifiers.

Method	Ages[15]	RVM [2]	QM [3]	Ages$_{lda}$ [15]	Ours	SVM [2]	RUN1[14]	LARR[5]	Ours with perfect classification
MAE	6.8	6.7	6.6	6.2	**6.2**	5.9	5.8	5.1	**3.4**

Table 3. MAE obtained by different methods. With perfect age group classification, we obtain the best result. However RUN1, LARR and warping velocity SVM [2] performs better than our overall hierarchical framework.

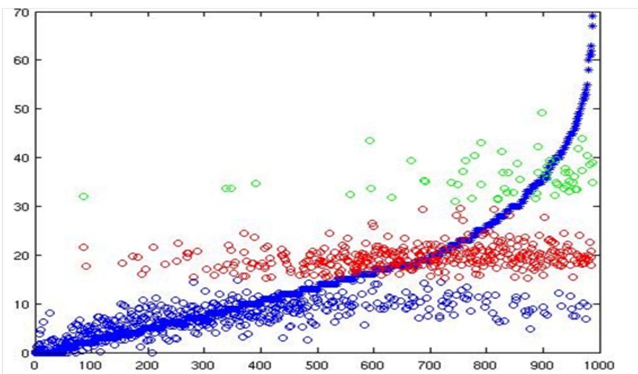

Fig. 4. The estimated age (by our hierarchical approach) of the test images are shown as circles and the ground truth age are shown as asterisks. The different colors represent classification into different age groups.

accuracy of the hierarchical approach, we would like to optimize the number and range of age groups. Also, we would like to incorporate texture features in our framework, which should improve the age estimation accuracy in the age group of 30 and above.

5. REFERENCES

[1] N. Ramanathan, R. Chellappa, and S. Biswas, "Computational methods for modeling facial aging: A survey," *Journal of Visual Languages and Computing*, vol. 20, June 2009.

[2] P. Turaga, S. Biswas, and R. Chellappa, "The role of geometry for age estimation," in *ICASSP*, 2010.

[3] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE TPAMI*, 2002.

[4] Y. Fu, Y. Xu, and T. S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," in *ICME*, 2007.

[5] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE TIP*, 2008.

[6] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, 2001.

[7] B. Scholkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Computation*, 2000.

[8] M. Barker and W. Rayens, "Partial least squares for discrimination," *Journal of Chemometrics*, 2003.

[9] R.O. Duda, P. E. Hart, and D.G. Stork, "Pattern classification," *John Wiley & Sons*, 2000.

[10] "Fg-net aging database, available at http://www.fgnet.rsunit.com/," .

[11] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape modelstheir training and application," *CVIU*, 1995.

[12] Kwon and Lobo, "Age classification from facial images," *CVIU*, 1999.

[13] W. B. Horng, C. P. Lee, and C. W. Chen, "Classification of age groups based on facial features," *Tamkang Journal of Science and Engineering*, 2001.

[14] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," in *ICCV*, 2007.

[15] X. Geng, Z. H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE TPAMI*, 2007.