

Gender Recognition & Age Prediction

Mr. Raghvendra, Prof.Sandeep Sahu

Master of Engineering Scholar
Computer Science and Engineering Department.
Shri Ram Institute of Technology.
Jabalpur (M.P.), India
patel.raghvendra@gmail.com

Head of CSE Department
Computer Science and Engineering Department.
Shri Ram Institute of Technology.
Jabalpur (M.P.), India
sandeep.sahu12@gmail.com

Abstract— over the recent years, a great deal of effort has been made to age estimation & gender recognition from face images. It has been reported that age can be accurately estimated under controlled environment such as frontal faces, no expression, and static lighting conditions. However, it is not straightforward to achieve the same accuracy level in real-world environment because of considerable variations in camera settings, facial poses, and illumination conditions. In this paper, we apply a recently-proposed machine learning technique called covariate shift adaptation to alleviating lighting condition change between laboratory and practical environment. Through real-world age estimation experiments, we demonstrate the usefulness of our proposed method.

Keywords— *Face Detection, Skin Color Segmentation, Face Features extraction, Features recognition, Fuzzy rules.*

I. INTRODUCTION

Human's faces reveal various information including gender, age and ethnicity. They provide important cues for many applications, such as biometric authentication and intelligent human-computer interface. In this paper, we present a new method that can identify humans' genders from their face images. In the past, many researches devote to finding good image features for gender recognition. Among them, Adaboost [1] [2] [3] [4] is a good tool for feature selection. There are many gender-recognition algorithms constructed based on AdaBoost. Wu and Ai [5] proposed a boosting method based on the look-up-table (LUT) weak classifiers. In their work, they train three different detectors (gender, ethnicity, and age) to acquire demographic information by using the two-class Real AdaBoost algorithm [2]. Shakhnarovich and Viola [6] used Viola and Jones' cascaded AdaBoost method [1] to train a face detector, which is a linear combination of the weak classifiers selected from rectangle features. They then selected a set of weak classifiers from the face detector for gender and ethnicity recognition. Instead of rectangle features, Shummet and Henry [7] designed a new set of weak classifiers which use the relationship between two pixels' intensities as features, and show that the recognition rate can be further improved. Instead of AdaBoost, Moghaddam and Yang [8] investigated the use of nonlinear support vector machines (SVMs) to classify gender from face images. With the Gaussian radial basis function (RBF) kernel, a high. However, the computational loads of these two approaches are high. They are thus not suitable for

real-time applications. Among the above, the Shakhnarovich and Viola's method [6] is probably the most efficient one in terms of the computational cost for real applications. It is because that some of the rectangle features used for face detection are re-used for gender recognition. Hence, in this method, we do not need to recompute the features once the face has been detected, and so the entire computational time (including both face detection and gender recognition) can be reduced. In addition, it is also well known that the evaluation of rectangle features can be considerably speeded up by using the integral-image technique [1]. However, the recognition rate of [6] still needs a considerable improvement. In this paper, we develop a fast gender recognition algorithm based on rectangle features too. Rectangle features can be used to describe sub-regions of a human face, and hence pixel-wise data can be transformed into component wise data. We cascade several rectangle features into a feature vector called RFV. The RFV is then served as a descriptor for faces to identify the gender. We did a comparative study by employing RFV in various kinds of classifiers, such as the nearest-neighbor (NN), principle component analysis (PCA), and nonlinear SVMs, and suggest an effective detector for Gender recognition. Unlike the method of Shakhnarovich and Viola [6], we do not restrict ourselves to select the rectangle features only from those used to construct the face detector. In other words, we allow the rectangle features to be selected arbitrarily in a large feature pool. In this way, the rectangle features selection can be more discriminative, and hence our approach is more accurate for gender recognition. Although in our approach, the rectangle features used for gender recognition and face detection may not

be the same, they still share the same integral image. Hence, we can still re-use the integral image, originally built for face detection, to efficiently evaluate the rectangle features for gender recognition. The total computational time (including both face detection and gender recognition) can thus be saved as well tools applied in this area will be described.

II. LITERATURE SURVEY

Identifying the age and gender information of a speaker given a short speech utterance is a challenging task and has gained significant attention recently. Metze et al. (2007) compared four approaches for age and gender recognition from telephone speech; these included a parallel phoneme recognizer system to compare the Viterbi decoding scores for each category-specific phoneme recognizer, a system using dynamic Bayesian networks to combine several prosodic features, a system based solely on linear prediction analysis, and a GMM system based on MFCCs. It was reported in Metze et al. (2007) that the parallel phone recognizer system performs as well as human listeners on long utterances but its performance degrades on short utterances while the system based on prosodic features, such as fundamental frequency (F0), jitter, shimmer and harmonics-to-noise-ratio, has shown relative robustness to the variation of the utterance duration. More recently, novel acoustic features (Ajmera and Burkhardt, 2008), frame and utterance based acoustic-prosodic joint features (Spiegel et al., 2009; Meinedo and Trancoso, 2010; Gajsek et al., 2010; Eyben et al., 2009), lexical features (Wolters et al., 2009) as well as fuzzy SVM modeling (Nguyen et al., 2010) have all been proposed to improve the recognition performance. In Ajmera and Burkhardt (2008), the discrete cosine transform is applied to the cepstral coefficients and the cepstral trajectories corresponding to lower (3-14 Hz) modulation frequencies provide best discrimination. Prosodic features (pitch, energy, formants, vocal tract length warping factor, speaking rate, etc) and their functional can also be added to the cepstral features at the frame or utterance level to enhance the performance (Spiegel et al., 2009; Meinedo and Trancoso, 2010; Gajsek et al., 2010; Eyben et al., 2009; Wolters et al., 2009). In addition to the prosodic features, novel lexical level features like word-class frequencies have also been proposed for age recognition purpose (Wolters et al., 2009). In the fuzzy SVM modeling method proposed by Nguyen et al. (2010), a fuzzy membership is assigned as a weight to each training data point to increase the robustness against noise and outliers. Furthermore, techniques from speaker verification and language identification applications such as GMM-SVM mean super vector systems (Bocklet et al., 2008), nuisance attribute projection (NAP) (Dobry et al., 2009), anchor models (Dobry et al., 2009; Kockmann et al., 2010) and Maximum-Mutual-Information (MMI) training (Kockmann et al., 2010) have been successfully applied to speaker age and gender identification tasks to enhance the performance of acoustic level modeling. In Dobry et al. (2009), anchor modeling utilizes a back end SVM to model the distribution of similarity scores between training data and all the anchor speaker models. Due to the different aspects of modeling, combining different classification methods together can often significantly improve the overall performance (Müller and Burkhardt, 2007; van Heerden et al., 2010; Meinedo and Trancoso, 2010; Bocklet et al., 2010; Kockmann et al., 2010; Lingenfeller et al., 2010). Humans perceive gender not only based on the face, but also on the surrounding context such as hair, clothing and skin tone [15, 6], gait [14] and the whole body [6, 1]. Below, we review relevant work on gender prediction from facial images only. The problem of gender

classification based on human faces has been extensively studied in the literature [20, 3]. There are two popular methods. The first one is proposed by Moghaddam et al. [20] where a Support Vector Machine (SVM) is utilized for gender classification based on thumbnail face images. The second was presented by Baluja et al. [3] who applied the Adaboost algorithm for gender prediction. Recently, due to the popularity of Local Binary Patterns (LBP) in face recognition applications [2], Yang et al. [30] used LBP histogram features for gender feature representation, and the Adaboost algorithm to learn the best local features for classification. Experiments were performed to predict age, gender and ethnicity from face images. A similar approach was proposed in [25]. Other local descriptors have also been adopted for gender classification. Wang et al. [27] proposed a novel gender recognition method using Scale Invariant Feature Transform (SIFT) descriptors and shape contexts. Once again, Adaboost was used to select features from face images and form a strong classifier. Gao et al. [10] performed face-based gender classification on consumer images acquired from a multi-ethnic face database. To overcome the non-uniformity of pose, expression, and illumination changes, they proposed the usage of Active Shape Models (ASM) to normalize facial texture. The work concluded that the consideration of ethnic factors can help improve gender classification accuracy in a multiethnic environment. A systematic overview on the topic of gender classification from face images can be found in [17]. Among all the descriptors that encode gender information such as LBP [25], SIFT [26] and HOG [6], the LBP has shown good discrimination capability while maintaining simplicity [17]. To establish a base-line for appearance based methods, we use LBP in combination with SVM to predict gender from facial images in this work. Although in previous work [22, 29] geometry features were used as a priori knowledge to help improve classification performance, none of the aforementioned approaches, unlike our work, focused explicitly and solely on facial metrology as a means for gender classification. Perhaps our work is more closely related to earlier research by Shi et al. [23, 24] on face recognition using geometric features, where they used ratio features computed from a few anatomical landmarks. However, we take a more comprehensive look at the explicit use of facial geometry in solving the problem of gender classification. We use solely metrological information based on landmarks, which may or may not be biologically meaningful. In our approach, the local information from independent landmarks is used instead of holistic information from all landmarks.

III. PROPOSED ALGORITHM

A. Gender Recognition

1. Input an Image
2. Detect skin area in Input Image.
3. Detect Features like eyes and mouth in skin region.
4. If Features detected then go to step 5 else step 1.
5. Crop Face.
6. Load Database Male Females features.
7. Locate Features in a face area.
8. Count Male & female Features.
9. Filter Counted features into strong & weak features.
10. Form Fuzzy Conclusion from features & Display gender result.

B. Age Prediction

1) Training

1. Select an Input Image.
2. Detect skin area in Input Image.
3. Detect Features like eyes and mouth in skin region.
4. If Features detected then go to step 5 else step
5. Crop Face.
6. Save Face into Database with its age.
7. Repeat step 1 to 6 for 100 images (Training Images)

2) Testing

1. Select an Input Image.
2. Detect skin area in Input Image.
3. Detect Features like eyes and mouth in skin region.
4. If Features detected then go to step 5 else step
5. Crop Face.
6. Load faces Images from training directory & Match With input face image.
7. Retrieve Match image age from database.
8. Display Result.
9. Stop

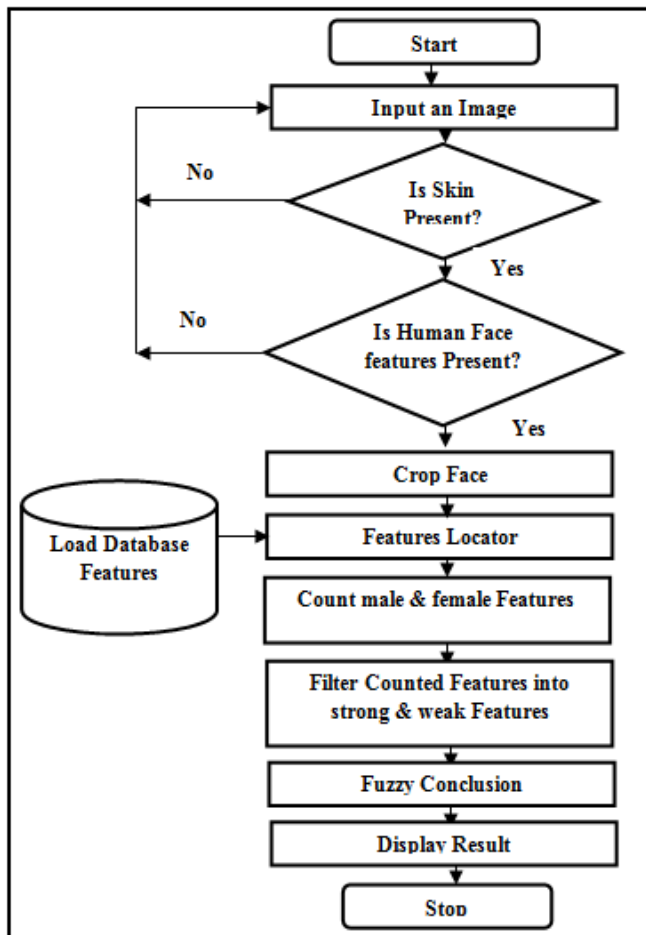


Fig. 1 Data Flow diagram of Gender Recognition

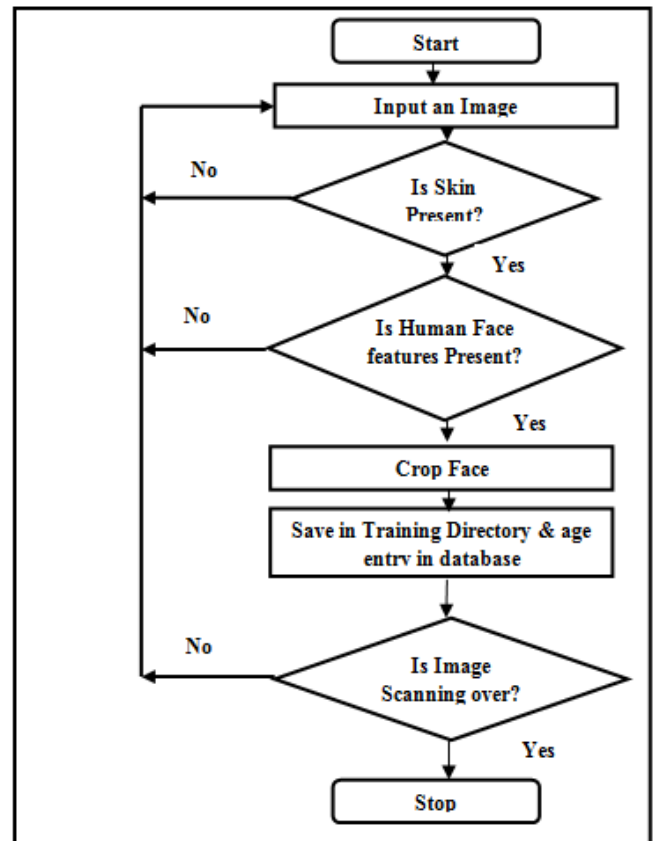


Fig. 2 Data Flow diagram of Gender age Prediction (Training)

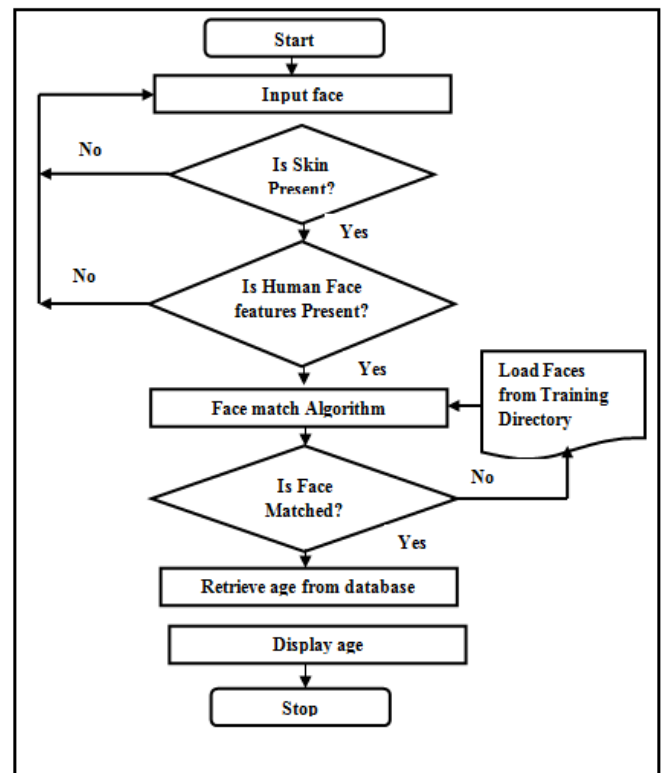


Fig. 3 Data Flow diagram of Gender age Prediction (Testing)

C. Face Detection

Skin color plays a vital role in differentiating human and non-human faces. From the study it is observe that skin color pixels have a decimal value in the range of 120 to 140. In this project, we used a trial and error method to locate skin color and non skin color pixels. But many of the times, system fails to detect whether an image contains human face or not (i.e. for those images where there is a skin color background).an image

is segmented into skin color and non-skin color pixels with the equations

$$180 \leq |Pxy| \leq 120 \quad \dots\dots\dots\text{eq. 1}$$

where $Pxy = \text{pixel at position } xy$

The skin pixels values are set to 1(i.e. #FFFF) and non skin pixels are set to 0(i.e. 0000). The pixels are collected and set as per equation

IF
 $\lim_{i \rightarrow 1} n (\int_1^3 120 \leq |Pxy| \leq 180) = 1 \quad \dots\dots\dots\text{eq 2}$

ELSE

$$\lim_{i \rightarrow 1} n (\int_1^3 180 \leq |Pxy| \leq 120) = 0 \quad \dots\dots\dots\text{eq 3}$$

where $n = \text{total number of pixels of input image}$

Literature review point out that, FACS system technique is based on face features extractions like eye, nose, mouth, etc. In this project, we minimize the number of features (i.e. only eyes and mouth) but given the more weight age for fuzzy rules formations from these extracted features. Face extractions consist of following steps

1. Let W and H are the width and height of skin and non-pixel image as shown in fig 5
2. Read the pixel at position (0,H/2) which is a middle of i.e. left side of image.
3. Travers a distance $D1 = W/6$ in horizontal direction to get the start boundary pixel of skin region.
4. Travers a distance $D2 = H/6$ from a pixel position (W/6, H/2) in upward directions. Same may do in downward direction and locate the points X1, X2.
5. Travers a distance $D3 = W/3$ from the point X1 and locate the point X3. Same do from the point x2 and locate the point X4.
6. Crop the square image as shown.

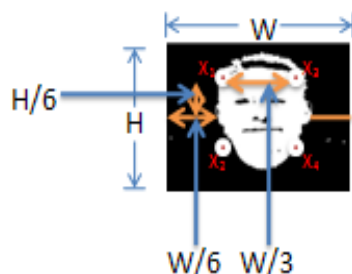


Fig. 4 Detected Face Areas



Fig. 5 Face Features Extraction

Human face is made up of eyes; nose, mouth and chine etc. there are differences in shape, size, and structure of these organs. So the faces are differs in thousands way. One of the common methods for face expression recognition is to extract the shape of eyes and mouth and then distinguish the faces by the distance and scale of these organs. The face feature extractions consist of following steps

1. Let W and H are width and height of an image shown in Fig 4
2. Mark pixel $P_i (W/2, H/2)$ as centre of image.
3. Travers a distance $H/8$ from the pixel P_i towards upward and mark a point K_1 .

4. Travers a distance $W/3$ from the point K_1 towards leftward and mark a point K_2 .
5. Travers a distance $H/10$ towards downward from the point K_2 and mark a point K_3 .
6. Travers a distance $W/4$ from the point K_3 towards right and mark the point K_4 .
7. Travers a distance $H/10$ from the point K_4 toward up and mark the point K_5 .
8. Same steps are repeated for extracting the right eye and mark the point $N_2, N_3, N_4,$ and N_5 .
9. Travers a distance $H/8$ from the point P_i towards downward and mark the point M_1 .
10. Travers a distance $W/6$ towards left and right from the point M_1 and marks the point M_2 and M_3 .
11. Start with the point M_2 traverse a distance $H/10$ towards downward and mark the point M_4 .
12. Travers a distance $W/6$ from the point M_4 towards right and mark the point M_5 . Same may do from point M_5 and mark the point M_6 .
13. Travers the distance $H/10$ from M_6 towards up that meets to the point M_3 .
14. See the below image.

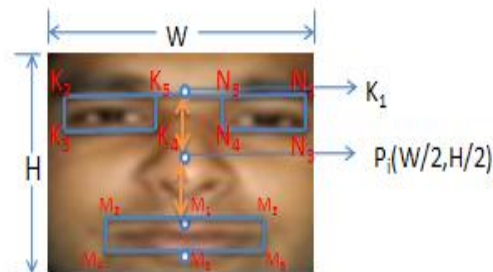


Fig. 6 Face crop

1. $\text{Dist } |P_i - K_1| = H/8$
2. $\text{Dist } |K_1 - K_2| = \text{Dist } |M_1 - M_2| = \text{Dist } |M_1 - M_3| = \text{Dist } |M_4 - M_5| = \text{Dist } |M_5 - M_6| = W/3$
3. $\text{Dist } |K_2 - K_3| = \text{Dist } |K_4 - K_5| = \text{Dist } |N_2 - N_3| = \text{Dist } |N_4 - N_5| = \text{Dist } |M_2 - M_4| = \text{Dist } |M_1 - M_5| = \text{Dist } |M_3 - M_6| = H/10$
4. $\text{Dist } |K_3 - K_4| = \text{Dist } |K_5 - K_2| = \text{Dist } |N_3 - N_4| = \text{Dist } |N_5 - N_2| = W/4$

D. Features Location

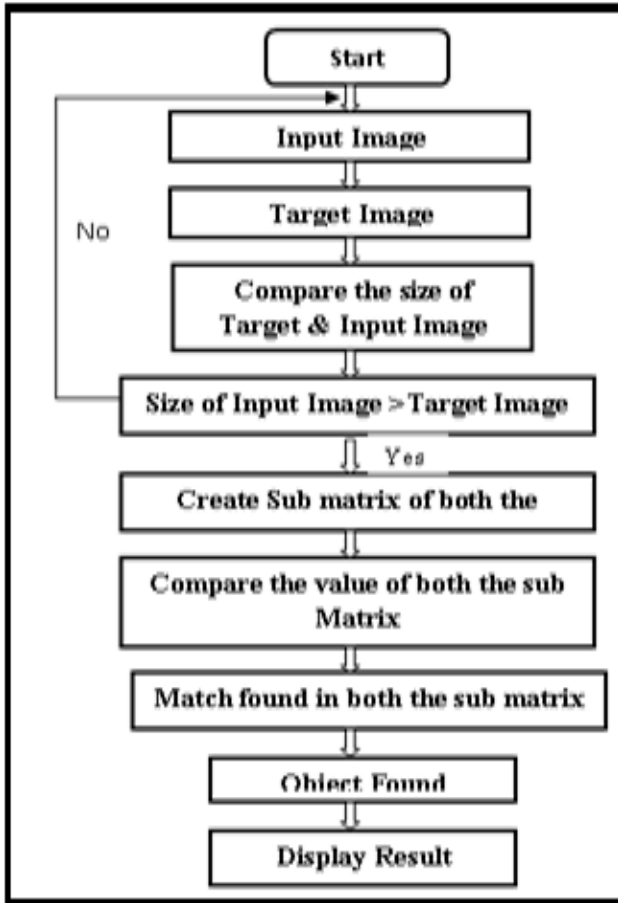


Fig.7 Data Flow diagram for locating features of training Database into input image face region

In this project we create a system that is useful to find the object that is present inside the image. In this we first take a image as an input for which we have to find the object. After selecting the input image we have to focus on the target image. After getting both input image and target image we compare the size of both the image. If the size of input image is greater than target image then we proceed with our system otherwise we are going to focus on the input image. If the above criteria get satisfied then we create the sub matrixes of both the images for example we create 3 X 3 matrixes. After creating the sub matrixes we compare both the sub matrixes of both the image. This Matrix can be created with the help of pixels present in both the images. If matching found then we concludes that the given object is found in the other image. The object is shown as an output by creating red boxes on that object. The output object image is displayed and the co ordinates of that object that is the height and width from top and bottom in the form of coordinates is shown with the help of im tool.

E. Fuzzy Rules & Conclusion

Table I Fuzzy Rules

Sr no	Male Features	Female Features	Fuzzy Conclusion
1	Beard		Male
2	Eyebrows		Male
3	-	Bindi	Female
4	-	Neckless	Female
5	-	Haar	Female
6	-	Ear ring	Female
7	Beard	Bindi	Male
8	Eyebrows	Bindi	Female
9	Beard	Bindi,Neckless	ambiguous
10	Bottleneck	Bindi,Neckless	female

We forms too many fuzzy rules that gives us appropriate result

Table II Proposed Methodology result

Sr No	Images tested	Gender Predicted	% of Gender prediction	Age Predicted	% of Age Prediction
1	5	5	100	4	80
2	10	7	70	6	60
3	15	12	80	10	66.66
4	20	16	80	14	70

1. Male & Female Features for gender recognition =50M/50F
2. Training directory faces for age prediction=100

IV. CONCLUSION

This paper presented an approach to age and gender classification from still image. We find that the proposed methodology system performs on average comparably to Human listeners, although the performance of our classifiers is worse on short utterances. A simple “majority voting” combination study did not improve classification accuracy, presumably due to the systematic nature of confusions. The result of a user study [7] shows that use patterns of proposed systems for senior citizens differ significantly from those of adults or young callers. Given that most confusion appear between neighboring classes, we believe the overall acceptance of proposed systems can be increased significantly by providing tailored versions of such systems, which adapt characteristics such as the degree of automation in a caller pre-selection scenario, order of presentation of options. In these scenarios, age and gender classification is not used to limit access (e.g. as in protection of minors), but to increase user satisfaction by providing individualized services even in the absence of knowledge about the caller’s identity.

References

1. Edward D. Mysak, “Pitch duration characteristics of older males,” Journal of Speech and Hearing Research, vol. 2, pp.46–54, 1959.
2. Sue E. Linville, Vocal Aging, Singular Publishing Group, SanDiego, CA; USA, 2001.
3. Christian M`uller, Frank Wittig, and J`org Baus, “Exploiting speech for recognizing elderly users to respond to their special needs,” in Proc. Euro speech 2003, Geneva; Switzerland, Sept.2003, ISCA.
4. Nobuaki Minematsu, Mariko Sekiguchi, and Keikichi Hirose, “Automatic estimation of one’s age with his/ her speech based upon acoustic modeling techniques of speakers,” in Proc.ICASSP 2002, Orlando, FL; USA, May 2002, IEEE.
5. Izhak Shafran, Michael Riley, and Mehryar Mohri, “Voice signatures,” in Proc. ASRU 2003, U.S. Virgin Islands, Dec. 2003, IEEE.
6. Susanne Sch`otz, “Automatic prediction of speaker age using CART,” Term paper for course in Forensic Phonetics, G`oteborg University

7. European Language Resources Association (ELRA), "<http://www.speechdat.org/>," <http://www.elra.info/>.
8. Jitendra Ajmera, "Effect of age and gender on LP smoothed spectral envelope," in Proc. Speaker Odyssey. 2006, IEEE.
9. Loredana Cerrato, Mauro Falcone, and Andrea Paoloni, "Subjective age estimation of telephonic voices," Speech Communication, vol. 31, no. 2–3, pp. 107–102, 2000.
10. The FG-NET Aging Database. <http://www.fgnet.rsunit.com/>.
11. Y. Fu, Y. Xu, and T. S. Huang. Estimating human age by manifold analysis of face pictures and regression on aging features. Proceedings of the IEEE Multimedia and Expo, pages 1383–1386, 2007.
12. G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang. A study on automatic age estimation using a large database. International Conference on Computer Vision in Kyoto (ICCV 2009), pages 1986–1991, 2009.
13. A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for no orthogonal problems. Technometrics, 12(3):55–67, 1970.
14. T. Kanamori, S. Hido, and M. Sugiyama. A least-squares approach to direct importance estimation. Journal of Machine Learning Research, 10:1391–1445, 2009.
15. K. J. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. Proceedings of the IEEE 7th International Conference on Automatic Face and Gesture Recognition (FGR 2006), pages 341–345, 2006.
16. B. Schölkopf and A. J. Smola. Learning with Kernels, MIT Press, Cambridge, MA, USA, 2002.
17. H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. Journal of Statistical Planning and Inference, 90(2):227–244, 2000.
18. M. Stone. Cross-validated choice and assessment of statistical predictions, Journal of the Royal Statistical Society, Series B, 36:111–147, 1974.
19. M. Sugiyama, M. Kawanabe, P. L. Chui. Dimensionality reduction for density ratio estimation in high-dimensional spaces. Neural Networks, 23(1):44–59, 2010.
20. M. Sugiyama, M. Krauledat, and K.-R. Müller. Covariate shift adaptation by importance weighted cross validation. Journal of Machine Learning Research, 8:985–1005, May 2007.
21. M. Sugiyama, M. Yamada, P. von Bülow, T. Suzuki, T. Kanamori, and M. Kawanabe. Direct density-ratio estimation with dimensionality reduction via least-squares hetero-distributional subspace search, Neural Networks.
22. F. H. C. Tivive and A. Bouzerdoum. A gender recognition system using shunting inhibitory convolutional neural networks. Proceedings of the International Joint Conference on Neural Networks (IJCNN '06), pages 5336–5341, 2006.
23. K. Ueki, M. Sugiyama, Y. Ihara. A semi-supervised approach to perceived age prediction from face images. IEICE Transactions on Information and Systems, to appear
24. K. Ueki, M. Miya, T. Ogawa, T. Kobayashi. Class distance weight locality preserving projection for automatic age estimation. Proceedings of IEEE International Conference on Biometrics: Theory, Applications and System (BTAS 2008), pages 1–5, 2008.
25. M. Yamada, M. Sugiyama, G. Wichern, and J. Simm. Direct importance estimation with a mixture of probabilistic principal component analyzers. IEICE Transactions on Information and Systems, E93-D(10), 2846–2849, 2010.