ISBN: 978-0-6480147-3-7

Real-Time Age and Gender Estimation from Face Images

Jang-Hee Yoo, So-Hee Park, and Yongjin Lee

ETRI (Electronics and Telecommunications Research Institute), Daejeon, S. Korea

Corresponding author's e-mail: jhy@etri.re.kr

Abstract: *In this paper, we describe an automated real-time system that estimates age and gender by utilizing a set of facial image sequences from a video camera. The age and gender estimation system consists of four steps: i) detection and extraction of the facial region from input video; ii) selection of the frontal face images from the extracted facial regions using head pose estimation; iii) duplicated face detection and removal by tracking the faces; and iv) age and gender estimation using statistical facial features. Here, LBP features with AdaBoost classifiers are used to detect the face region in a video frame, and the frontal face images are selected using a 3D pose estimation method. In addition, a particle filter-based tracking framework is employed to remove duplicated faces and to improve the accuracy of people counting, and Gabor-LBP features are used to estimate age and gender using a linear SVM and Adaboost classifiers. In experiments, a large number of face datasets are used to train and evaluate the proposed method, and higher performance is achieved in terms of age and gender estimation: 72.53% for age and 98.90% for gender.*

Keywords: Face Attributes, Age and Gender Estimation, Face Recognition.

1. INTRODUCTION

Studies of face recognition have increased due to the large number of application fields, such as user authentication, targeted advertisements, video surveillance and human-robot interaction. Various methods have been developed over the past two decades, and the most widely used methods depend on holistic representation approaches (Bartlett et al., 2002; Etemad et al., 1997; Truk et al., 1991), such as PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis), and ICA (Independent Component Analysis). These approaches are very simple and effective for frontal faces, but are not robust in terms of pose, illumination and expression changes. According to the recent literature, non-statistical learning methods such as Gabor wavelets and LBP (Local Binary Patterns) have been successfully applied in face recognition applications (Gao et al., 2008; Serrano et al., 2010).

On the other hand, age and gender estimation from facial images has also received increasing attention, and many methods have been developed. One major issue in age and gender estimation is how to extract effective representation features from a facial image (Chen et al., 2017). A comprehensive survey of methods and data can be found in (Fu et al., 2010; Han et al., 2015). More recently, deep learning approaches such as CNN (Convolutional Neural Networks) have improved the performance of face recognition and age/gender estimation (Chen et al., 2017). As a commercial application, NEC and Microsoft have produced one of the most practical age and gender estimation systems; semi-supervised learning is employed for age estimation by NEC and deep neural networks by Microsoft.

In this paper, we propose an automated real-time system for age and gender estimation from face images. To achieve this, face detection and pose estimation methods are adopted to acquire frontal face images. Then, LBP and Gabor filters are used to extract face features, and a supervised learning approach using Adaboost is employed to estimate age and gender. To validate the performance of the proposed method, experiments are conducted using a benchmark database. The rest of this paper is organized as follows. Section 2 describes the proposed age and gender estimation system. Experimental results are described in Section 3 which evaluate the proposed method against the benchmark database. Finally, Section 4 concludes this paper, with remarks on future work.

2. REAL-TIME AGE AND GENDER ESTIMATION SYSTEM

Automated age and gender estimation of the human face involves detecting, tracking and normalizing the face in an image sequence. The features extracted from the face image are used to estimate age and gender, and a face tracking method is employed to measure stay time and to count people. Figure 1 illustrates the system architecture used in this study.

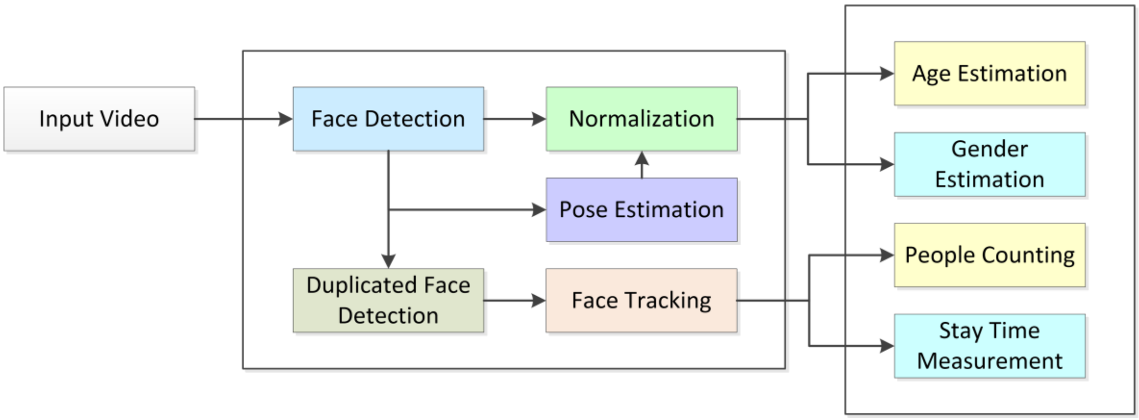

Figure 1. System architecture for age and gender estimation

2.1 Face Detection and Pose Estimation

The LBP-AdaBoost approach (Paris et al., 2011) is used for face detection and eye localization in the input frames. This method adopts the LBP histogram as feature vectors, and the two-class LDA as weak classifiers. LBP features are invariant to monotonic changes in illumination and are computationally efficient. The process of LBP-AdaBoost for face and eye detection is shown in Figure 2. In order to reduce processing time, face detection is performed for only every third input frame, and 64×64 is the minimum size of the face region used for detection. Facial images using localized eye positions are aligned to apply the face estimation algorithm (Park et al., 2011). In addition, an automatic and monocular head pose estimation method is used for acquiring frontal faces using pitch, yaw and roll values. There are many head pose estimation algorithms; however, a point-based algorithm allows fast estimation with no temporal information (Kim et al., 2011). Thus, POSIT (Pose from Orthography and Scaling with ITerations), a hybrid pose estimation algorithm based on both algebraic and optimizing algorithms, is applied for head pose estimation. POSIT has advantages from both approaches and has good speed and accuracy (Kim et al., 2011).

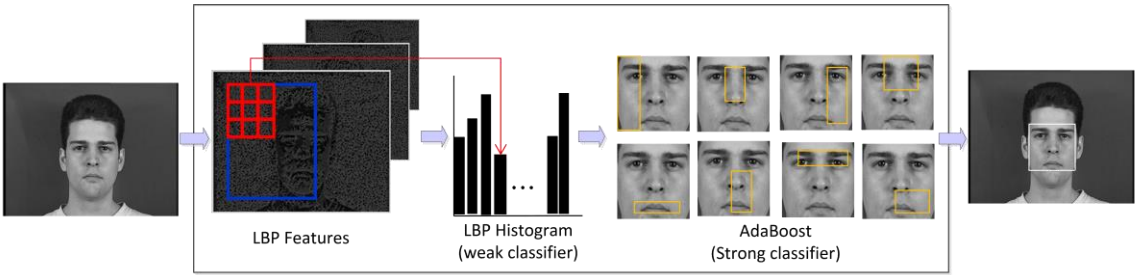

Figure 2. LBP-AdaBoost approach for face detection

2.2 Face Tracking and People Counting

The people counting method based on face tracking consists of face detection, face tracking, tracking failure detection, data association and counting modules. The tracking framework is based on a multi-person tracking algorithm (Choi et al., 2015). An HSV (Hue Saturation Value) histogram is employed as an observation model, and Gaussian perturbation is employed as a motion model in a particle filter-based tracking framework. In order to detect tracking failure, we define a criterion for tracking loss or success as:

$$\text{Status}(k) = \begin{cases} 0 & \text{if } \rho[l_{\hat{\mathbf{s}}_t^k}, l_t^k] < \rho_{\text{th}}, \\ 0 & \text{if } \rho[l_{\hat{\mathbf{s}}_t^k}, l_t^k] - \rho[l_{\hat{\mathbf{s}}_{t-1}^k}, l_{t-1}^k] < \Delta_{\text{th}}, \\ 1 & \text{otherwise,} \end{cases} \tag{1}$$

where $l_{\hat{s}t}{}^k$ is the colour histogram constructed at the estimated position of the *k*th face, $l_t{}^k$ is the target colour histogram of the *k*th face, and ρ_{th} and Δ_{th} are predefined threshold values. If a tracking failure is detected by Equation 1, the lost signal (zero) will be transmitted to the face detection module to request a redetection of the lost face. Otherwise, the tracker estimates the state of the current tracking face using particle filtering.

To assign the face detection result to the trackers, a data association method is required. Greedy data association is used to solve the assignment problem, with the matching score defined as:

$$M(\mathbf{s}_{\text{tr}}, \mathbf{s}_{\text{det}}) = \alpha\, p_N(\mathbf{s}_{\text{tr}} - \mathbf{s}_{\text{det}}) + \rho[l_{\text{tr}}, l_{\text{det}}], \tag{2}$$

where $p_N(\mathrm{s}_{\text{tr}} - \mathrm{s}_{\text{det}})$ is a Gaussian distribution with a mean of zero and standard deviation σ_d, and α is a weighting factor for the distribution. The face detection result with maximum matching score is associated with the tracker. The associated face detection result and the tracker are then excluded from the matching pool, and this procedure is repeated until no valid pair remains in the matching pool. The proposed method counts the people crossing a pre-configured line with a pre-configured counting direction (e.g. top to bottom, left to right etc.).

2.3 Age and Gender Estimation

After the face and eyes are detected, the facial image can be normalized as a fixed-size image using the localized eye positions. Then, the Gabor-LBP histogram framework (Gao et al., 2008) is used to extract the features for representation power of the spatial histogram (Lee et al., 2012). Using Gaussian derivative filters, the dimensions of the facial features can be reduced by 60% compared with the Gabor wavelet filters. The derived first-order and second-order Gaussian derivative filters are shown in Figure 3. As shown in the figure, the multiple Gaussian derivative filter images are converted by convolving the normalized facial images with the Gaussian derivative filters. Then, each Gaussian derivative filter image is transformed into an LBP map image using the LBP operator, as shown in Figure 3. The LBP patterns at each pixel can be completed by summing the threshold values weighted by a power of two, which characterizes the spatial structure of the local image texture. Finally, each LBP map image is divided into non-overlapping sub-regions with a predefined bin size, and a total of 36 bins are used for each LBP map image. The LBP histograms of the sub-regions are concatenated to form the final histogram sequence as the face representation.

In general, age estimation can be characterized as a multi-class or regression problem, and gender estimation is defined as a two-class problem. Thus, although we can apply AdaBoost directly to gender estimation, applying Adaboost to age estimation is not straightforward. Figure 4 shows a hierarchical approach for age estimation. For example, suppose that we have eight age classes. Then, the training data can be divided into two sub-groups recursively, and Adaboost can be trained such that the two sub-groups are well-discriminated. As shown in the figure, a total of seven classifiers are required to train age data for estimation (Lee et al., 2016).

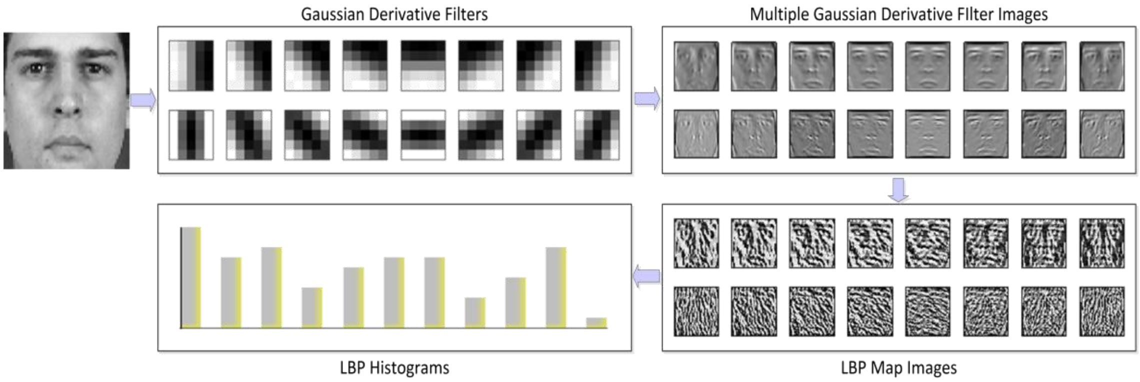

Figure 3. Face features using LBP with Gaussian derivative filters

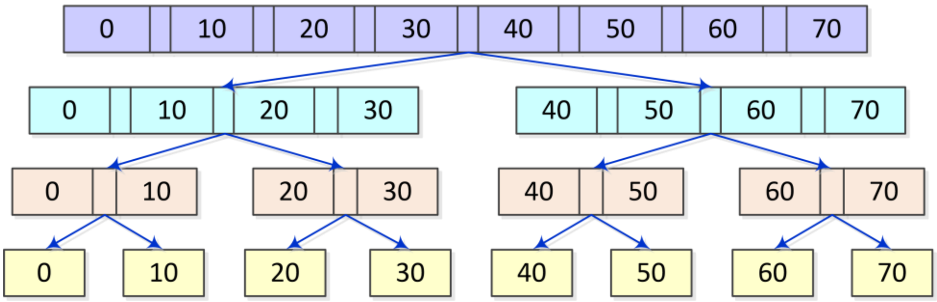

Figure 4. Hierarchical multi-classifier

3. EXPERIMENTAL RESULTS

In the experiments, 25,866 face images (21,392 for training and 4,474 for testing) are used for gender estimation. To estimate gender, Adaboost is trained until the classification for training data reaches nearly 100%. Table 1 shows the test results for the estimation accuracy of testing data. As can be seen from the table, the female dataset shows very good accuracy compared to the male dataset. The reason for this is that many images in the testing data were collected via internet surfing, and internet images are frequently smoothed using image processing tools such as Photoshop. Human features are very sensitive to texture and skin tone, and most of the selected features are located around the meaningful areas for recognition, such as eyebrows, nose, cheekbones and jaw-line. Nevertheless, we obtain a good performance of 98.90% CCR (Correct Classification Rate) in a real-time test.

Table 1. Gender estimation using facial images

Gender	Training (# of Subjects)	Testing (# of Subjects)	Test Result (CCR, %)
Male	12,525	2,092	78.63
Female	8,867	2,382	98.05
Total	21,392	4,474	87.37

We have much less training data for age estimation than for gender classification, and the gathering process of ground truth for age is rather difficult; even humans cannot estimate real age accurately. Thus, the estimation accuracy is low.

Table 2. Age estimation using facial images

Age	0-2	2-7	8-13	14-19	20-29	30-39	40-49	50-59	60-69	70+	Total
Training (# of Sub.)	33	59	114	430	500	500	500	500	239	90	2,965
Testing (# of Sub.)	34	38	33	154	118	72	104	49	30	9	641
Test Results (CCR, %)	85.29	68.42	57.58	61.69	56.78	37.50	44.34	53.06	63.33	77.78	56.32

Table 2 shows the number of experimental datasets and the accuracy for age estimation of the testing data. As can be seen from the table, similar classes are often confused with each other. For example, small children (2~7) are confused with older children (8~13), older children with teenagers (14~19), and thirties (30~39) with forties (40~49). Although the estimation performance is low, the results are quite meaningful. In fact, humans are also easily confused between these age groups. In addition, we obtained 72.53% CCR with a tolerance of ± 5 years, in a real-time test.

4. CONCLUSIONS

We describe a real-time age and gender estimation system with an AdaBoost classifier for facial images. The age and gender estimation system consists of face detection, pose estimation, face tracking and facial feature extraction modules. To estimate age and gender, the extracted facial features are used to train the AdaBoost classifier; people counting and stay time measurement are completed using duplicate face detection with the face tracking method. In a real-time test, estimation rates of 98.90% CCR for gender and 72.53% CCR for age are achieved. These results show promising performance and higher estimation rates than those of an earlier age and gender estimation approach.

ACKNOWLEDGMENTS

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00046, Basic Technology for Extracting High-level Information from Multiple Sources Data base on Intelligent Analysis).

REFERENCES

Bartlett MS, Movellan JR, and Sejnowski TJ (2002). Face Recognition by Independent Component Analysis, IEEE Trans. on Neural Networks, 13(6), pp.1450-1464.

Chen S, Zhang C, Dong M, Le J, and Rao M (2017). Using Ranking-CNN for Age Estimation, in Proc. of the Int. Conf. on Computer Vision and Pattern Recognition (CVPR), pp.742-750.

Choi JW, Moon DS, and Yoo JH (2015). Robust Multi-Person Tracking for Real-Time Intelligent Video Surveillance, ETRI Journal, 37(3), pp.551-561.

Etemad K and Chellappa R (1997). Discriminant Analysis for Recognition of Human Face Images, Journal of the Optical Society of America A: Optics and Image Science, and Vision, 14(8), pp.1724-1733.

Fu Y, Guo G, and Huang TS (2010). Age Synthesis and Estimation via Faces: A Survey, IEEE Trans. Pattern Anal. Intell. 32(11), pp.1955-1976.

Gao T and He M (2008). A Novel Face Description by Local Multi-channel Gabor Histogram Sequence Binary Pattern, in Proc. of the Int. Conf. on Audio, Language and Image Processing, pp.1240-1244.

Han H, Otto C, Liu X, and Jain AK (2015). Demographic Estimation from Face Images: Human vs. Machine Performance, IEEE Trans. Pattern Anal. Intell. 37(6), pp.1148-1161.

Kim WW, Park SH, Hwang JK, and Lee SY (2011). Automatic Head Pose Estimation from a Single Camera using Projective Geometry, in Proc. of the Int. Conf. on Info. Comm. and Signal Processing.

Lee HS, Chung YS, Yoo JH, and Won CH (2012). Face Recognition based on Sparse Representation Classifier with Gabor-Edge Components Histogram, in Proc. SITIS, pp.105-109.

Lee Y, Yoo JH, and Kim KH (2016). Feature Analysis for Human Gender and Age Recognition, Draft.

Paris S, Glotin H, and Zhao Z (2011). Real-time Face Detection using Integral Histogram of Multi-scale Local Binary Patterns, LNCS 6838, Springer-Verlag, pp.276-281.

Park SH, Yoo JH (2012). Robust Eye Localization using Correlation Filter, in Proc. ICCCS, pp.137-140.

Turk MA and Pentland AP (1991). Face Recognition using Eigenfaces, in Proc. CVPR, pp.586-591.

Serrano A, Diego IM, Conde C, and Cabello E (2010). Recent Advances in Face Biometrics with Gabor Wavelets: A Review, Pattern Recognition Letters, 31, pp.372-381.