Ethnicity Classification From Face Images, Literature Review

Zahraa Shahad Marzoog¹, Dr. Manal Hussein Nawir², Dr. Manal Hussein Nawir², Fatima Al Zegair³

¹School of engineering, University of Kerbala, Kerbala, Iraq.
²School of Engineering, Kerbala University, Kerbala, Iraq.
³School of Information Technology and Electrical Engineering, University of Queensland, Brisbane, Australia.

Abstract

Race classification has been a long-term challenge in the field of face recognition recently. As it is a key-demographic trait of individuals, it has been employed in real-world applications; for example, surveillance videos and online advertisement, Human-Computer Interaction, law enforcement, and demographic and biometric research. Consequently, different approaches have been suggested to tackle the race classification problems. This paper presents a literature review of state-of-art research that address the ethnic classification problem.

Keywords

Race classification, ethnicity classification, Deep learning, Convolution Neural Network.

Literature Review

Different methods have been proposed in race-based face recognition fields recently. Many proposed approaches have concentrated on the full-face picture when designing their models. These approaches have adopted mathematical and statistic techniques to extract discriminating features (Al-Humaidan and Prince 2021). For example, Lu et al. have suggested an ethnicity classification technique that examined the face at various scales (Lu, Jain, and others 2004). This method applied the Linear Disciminate Analysis (LDA) to facial images to enhance the classification performance. This adopted method has utilized 2630 sample faces of 263 subjects. Although this method demonstrates 96.3% accuracy, the database is classified into two classes only (Asian and non-Asian). Another model proposed by Guo et al. depended on the canonical correlation analysis (CCA) technique to broadcast ethnicity, gender, and age (Guo and Mu 2014). Manesh et al. have adopted a method that uses the Golden ratio mask by employing decision-making rules on different face zones, which are automatically separated (Manesh, Ghahramani,
and Tan 2010). Then, the SVM technique classified the extracted Gober features. Similar to Lu’s approach, this method has classified the database into Asian and non-Asian. Some methods have been based on skin or hair color to recognize ethnicity. In (Demirkus, Garg, and Guler 2010), two algorithms have been proposed, which are pixel intensity values and the Biologically Inspired Model (BIM) to classify ethnicity based on hair and face color. SVM classifier has been used to classify 600 images into three ethnic groups (Asian, African, and Caucasian) and achieved a better performance of 81.3%. Becerra-Riera et al. proposed a method that has codified information about different face regions with appearance and geometric features such as color and facial features’ shape (Becerra-Riera et al. 2018). Two feature combination strategies have been used with both RF and SVM as classifiers. The model classified the face images from the FERET dataset into three groups: white, black, and Asian, and achieved an accuracy of 93.7%. Wu et al. proposed a Look-Up-Table week classifier that depended on boosting method to extract demographic features (such as race and gender) from human faces (Wu, Ai, and Huang 2004). The proposed method comprises of three modules, including facial region detection, facial feature extraction, and demographic classification. The model has been tested on a dataset that includes three categories: 1771 Africans, 2306 Caucasians, and 2411 Mongoloids, and achieved accuracy of 95.0%, 96.1%, 93.5% in classifying Mongoloids, Caucasians, and Africans, respectively. Muhammad et al. proposed a race-based classification method based on two local descriptors’ types: Weber Local Descriptor (WLD) and Local Binary Pattern (LBP) (Muhammad et al. 2012). This model also utilized different distance classifiers, including Chi-square, Euclidean, and City-block. The result demonstrated that fusing WLD with LBP and utilizing City-block as a distance classifier performs better accuracy.

Some methods have addressed race recognition problems by focusing on a particular region in the face. For example, Lyle et al. proposed a method that selected texture features from the periorcular region by applying LBP feature-based technique to distinguish between Asians and non-Asians (Lyle et al. 2012). Sun et al. proposed a classification technique dependent on a hierarchal visual codebook to classify iris images. Although the method achieved a promising performance, its deficiency is evident because of the difficulty to extract iris texture due to the long-distance camera in practical applications. Xie et al. suggested a novel ethnic classification method that incorporated facial color-dependent features and Kernel class-based features to identify the race of the database, including Asian, African-American, and Caucasian (Xie, Luu, and Savvides 2012). This method has focused on the upper part of the face, including the periorcular region. Hosoi et al. suggested a new approach that cooperates retina sampling and the Gabor wavelet features to be classified using Support Vector Machines (SVM) technique. This approach has achieved 93%, 94%, and 96% in classifying Asian, African, and European, respectively, while there has been an issue with other ethnicities classification. Roomi et al. suggested a face detection method that used the Viola-Jones algorithm (Roomi et al. 2011). The authors have focused on extracting features from different regions, including lip color, skin color, and forehead area. The FERET, Yale dataset of Negroid, Mongolian, and Caucasian samples have been used in this method,
with an accuracy of 81% when classifying them. Marzoog et al. proposed a novel race classification model based on the geodesic distance technique as a feature extractor (Marzoog, Hasan, and Abbas 2022). Principal component analysis (PCA) has been utilized as a dimensionality reduction to decrease the dimension of the extracted features without losing information before classification. Support vector machine (SVM) and K nearest neighbour (KNN) have been used as classifiers. This method achieved an accuracy of 100% in classifying FERET DATASET into three ethnic groups: Asian, African and Caucasian.

Some suggested methods have recently concentrated on deep learning to solve the issue of race classification. Zhang et al. have exposed the deep CNNs technique in building a feature learning model that achieved promising performance in classifying halftone images (Zhang, Zhang, and Chen 2016). Wang et al. tackle race classification’s problem using a deep convolution neural network. This proposed model has utilized the dataset which is a combination of both public and self-collected datasets (Wang, He, and Zhao 2016). This method has achieved two binary classifications: black and white, and Chinese and Non-Chinese. It also attained a three-class classification: Uyghurs, Hung, and Non-Chines. Another CNN model named Hypotheses CNN pooling, proposed by Wei et al., to classify multi-labeled images classification. This work has used a random object segment of Hypotheses’ number as a framework’s input. Then, shared CNN connected to these hypotheses (Wei et al. 2015). Khan et al. suggested an approach that utilized CNN to extract deep features from seven different classes and created probability maps for them. The probabilistic classification method has been applied to various datasets, such as CAS-PEAL, VNHashes, and VMER, and achieved 99.2%, 92%, and 93.2%, respectively (Khan et al. 2013). Biag et al. proposed a method that used a convolution neural network to classify faces. Significant features, such as surface, color, and skin, were integrated with several secondary features to improve race classification images into Asian and non-Asian (Baig et al. 2019). The accuracy of this method, however, was 84.91%. Vo et al. suggested two race recognition frameworks: race recognition depended on CNN (RRCNN) and race recognition depended on deep learning architecture VGG (RRVGG) (Vo, Nguyen, and Le 2018). Those models achieved similar performance (88.64% and 88.87%, respectively). Greco et al. [21] proposed a VMER dataset and utilized four well-known convolution neural networks (ResNet 50, MobileNet V2, VGG-16, and VGG-face) (Greco et al. 2020). The best classification result was 94.1%. Heng et al. proposed a hybrid classification model by combining the classification result of CNN and the image ranking engine to benefit from features fitting between query and images in the dataset (Heng, Dipu, and Yap 2018). The hybrid feature vectors have been classified using SVM and attained an accuracy of 95.2%. Afifi et al. proposed a new method that combined holistic and local features of the face image (Afifi and Abdelhamed 2019). Four convolution neural networks have been used to classify the extracted features. An AdaBoost-based score fusion technique has been applied to assemble the scores obtained from CNN’s. This model has achieved accuracy ranging between 90.5% and 99.9% on four different datasets. Ahmed et al. suggested a new convolution neural network to address race recognition’s problem. This CNN model
includes nine layers, with the first six layers being the convolution layers. The last three layers are fully connected layers. CFD and UTK datasets have been utilized to test this model and classified the dataset into four ethnic groups: Asian, Caucasians, African and Indian. The model achieved an accuracy of 97%. Mohammad et al. suggested a new approach to support the security of the mobile environment based on ethnicity biometrics (Mohammad and Al-Ani 2018). This study proposed six Convolutional Neural Network models with different architectures to classify five ethnic groups: middle-east, Asian, African, white, and Hispanic. Before classification, the ocular region from face images of the FERET dataset has been extracted. Model-02 achieved the best accuracy of 98.35%, while Model-5 attained the lowest accuracy of 88.35%. Masood et al. utilized two strategies, namely artificial neural network and convolution neural network VGGNet, to classify the FERET dataset into three ethnic groups: Negro, Caucasian, and Mongolian (Masood et al. 2018). This method focused on classifying the extracted color characteristics and geometric features from face images. The accuracy results when using ANN and CNN were 82.4% and 98.6%, respectively.

**Conclusion**

This paper shows different research that has attempted to improve the robust race and classification models. Generally, most research has examined the race-discriminative features in human faces by integrating strategies for understanding, processing, and extracting information from facial images. Various research has addressed race classification by either using a whole face image as a one-dimensional feature or using a specific region of the face. Other recent research has focused on CNN to extract and classify face images into different ethnic groups.

**References**


image processing and graphics.


