Survey on Facial Feature Detection Methods in Offline Recommendation based Systems

KJ Harikrishnan[1], Akhil Mathew[2], Athul Nair[3]

[1,2,3]UG Scholar, Viswajyothi College of Engg and Technology, Ernakulam, Kerala

---

Abstract - *This paper analyses the utilization of facial feature extraction as a means of input for recommendation algorithms. Age and gender detection are two base level techniques that can be used to predict user interests as far as a recommendation algorithm is concerned. This paper discusses the utilization of such a recommendation algorithm in a non-specific context, except for assuming that the system under study has a visual output display, and camera-based input. The paper also introduces the concept of emotion detection as a means of obtaining feedback to improve the parameters dictating the process of recommendation for varying input values of age and gender. The paper aims to serve as a preliminary discussion of the state-of-the-art of using automated facial feature extraction based recommendation for intelligent display systems. The findings of this survey can be implemented in multiple scenarios where an intuitive display is applicable.*

Keywords: Image Processing, Facial Analysis, Human Computer Interaction, Recommendation, Computer Vision

1. INTRODUCTION

Organizations such as Facebook, Amazon, Netflix and Google utilize recommendation algorithms to provide personalized results to consumers. Facebook, for example, tracks posts and videos viewed by its clients to curate a specific selection of content on user feeds. It also utilizes this understanding of user interests to decide which advertisements should be displayed to which customer (based on likelihood of a user responding positively (i.e. "clicking") to it). A similar method is also used by Google Ads, which tracks user search history to know what the Google user is interested in. Controversially, both Google and Facebook have been accused of using personal information of consumers on their respective platforms (such as age, location, education/career status) to leverage their ad services. This is seen as unethical as both these corporations (among many others) are using user data without permission in order to maximize their own profit. Amazon, like many online retailers, has two methods to hone their product recommendation: one is to analyze each customer's past purchases/product views, and another is to analyze product associations (i.e. what products are often bought together). Media hosting sites, like Netflix or Spotify, recommend entertainment options based on individual viewing history and current popularity of content.[7]

A casual observer of these practices can readily note that essentially users are being categorized based on some data (the input to the algorithm), and some content is being prescribed based on what how the algorithm labels/clusters them (the output of the algorithm). [13] It is easy to obtain all this information about users on online platforms; however, how can one replicate a user-influenced recommendation in a physical, real world scenario? What kind of data could be used as input to a recommendation algorithm for such a system? How would such a system compare to its virtual counterpart?

2. PROBLEM STATEMENT

The paper aims to devise a system that will enable a recommendation system to be implemented in a physical, real life scenario. As mentioned in the Introduction, online applications have access to user history, and personal details of user, and thus these data can be used to classify the user and provide them suitable content that they may respond positively to. However, many limitations exist when attempting to implement this in the real world. For example, for an advertisement display in a mall/store, there is no guarantee that a system has access to a user's past purchases, or even their personal details. Data security and privacy are also big concerns: many consumers in the general public may not be comfortable knowing that their personal information can be viewed and analyzed by third party software. Thus, it is critical to opt for user data that can be extracted in real time, will not violate user privacy and yet give enough information about a user to classify them to a rudimentary extent and provide them with relevant content. The paper is solely focused on the data that can be used as input to such an algorithm, along with how to extract this information in real-time and does not delve into the specifics used to construct the algorithm itself.

3. SCOPE

An offline recommendation system that extracts user information in real-time and categorically provides predictive suggestions has a multitude of possible applications. It can be used, as exemplified in the

Problem Statement, to enhance the effectiveness of physical advertising by making it dynamic and customized. It is most effective in a commercial setting, as it replicates one of the most efficient and successful aspects of online commerce in a physical incarnation: targeted marketing. It can also be used in movie theaters and recreational parks like Disney Land to recommend demographic group-specific attractions. Government bodies could also use such displays in their offices to inform the public about campaigns or initiatives that they may be interested in.

4. ASSUMPTIONS

As the paper discusses various methods for obtaining user-related information to provide recommendations in a physical real time system, it is necessary to state the constraints that the system has to align with in order for the proposed techniques to be functional. For example, the authors have assumed that the system hardware has a camera facility of sufficient image quality, which allow for higher feature detection and extraction. The system is also presumed to have sufficient hardware and processing capabilities to execute image processing algorithms. The network speed should be sufficient to allow fast server to UI data transfer. To simplify the proposal, the researchers have assumed that the system has no access to any historical data (i.e. past selections made by the same user) that will influence future recommendations in any way. Thus, the accuracy and relevance of the recommendation is based solely on the visual information captured through facial processing of user, the result categorization, and categorical recommendation that is preset by system managers.

The methods are discussed based on utility of the data they provide (i.e. how they aid the algorithm), and in chronological order.

5. USER CATEGORIZATION

The specification given in the Problem Statement is that the input to the recommendation algorithm should be data that can be obtained in real time, isn't too personal, and yet provides sufficient background about the person so that a somewhat accurate and appropriate suggestion can be generated.

Demographic groups are one way of sorting people. In most of the potential use cases for this recommendation system, the areas of interest and types of content preferred would vary to some degree among the different demographic groups. Usually, demographic groups are defined by four essential features: age, gender, ethnicity and economic status. As economic status is a private piece of information, and ethnicity can't be obtained very accurately in a single image, real time identification system (also, accusations of stereotyping and racism are highly likely if a system utilizes ethnic categorization), the paper focuses on age and gender. Thus, the researchers shall explore various techniques to extract age and gender information from user images and submit this to the recommendation algorithm for labeling purposes. The actual content associated with each label shall be preset by a human entity, and in this discussion will be static.

Basically, all age and gender detection algorithms have three steps in common: face detection, feature extraction, gender/age labeling. Facial features used for the identification fall into three categories: global, local or hybrid. Global features are those that change from birth to adulthood (examples: size of cranium, distances between features). Local features are those that changes from adulthood to old age (examples: wrinkled, facial blemishes, hair growth patterns). Hybrid is a mix of global and local. There are several different facial representation models used for age recognition such Anthropometric Model, Active Appearance Model, Aging Pattern Subspace and Age Manifold. Either classification or regression can be used for Estimator learning algorithms. In the approach that utilizes classification, different age groups (example 0-10, 11-20, etc) are predetermined and the current subject is assigned to one of them based on analysis of features. In regression, however, an exact numerical value is assigned.[1]

Bekhouche et al propose an approach for gender classification based on Multi-Level Local Phase Quantization (ML-LPQ) features extracted from facial images.[5] An SVM model is used to predict the gender. The experiment was carried out on Groups database. For the experiment conducted, an accuracy of 79.1 % was obtained in case of gender classification. The approach delivered a better performance when compared with other Local Phase Quantization methods. However, feature selection has not been considered here, which can boost the overall performance of the approach.

Syed Musa Ali and his fellow researchers introduce a method to divide the facial images into 3 age groups based on biometric ratios and wrinkle analysis. [4]The process first detects facial characteristic using Haar-like features and then derives the biometric ratios. The highlighting feature of this method includes the fact that the methodology is independent of the type of classifier used as well as the size of the image. The data from the feature extraction phase helps to create biometric ratios using biometric distances between various facial points. Wrinkle density calculation is performed on the (binary) output of a Canny detection performed on a grayscale version of the image. The derived features are then fed to various machine learning algorithms such as the K-NN classifier, ID3 classifier, or Naïve Bayesian classifier. The system was tested on a large enough database showed a 20% increase in accuracy when wrinkles were introduced; however misclassification occurs when the forehead is covered with hair.

Sahib Khan and team preferred a method to analyze the age of human using wrinkle analysis as well.[11]However, they utilize Discrete Cosine Transform (DCT) , concentrating on the region of the cheeks and utilizing wrinkles analysis. The deciding factor here is based on the fact that as density of wrinkles or age(assuming greater age implies greater wrinkle density)the energy in higher coefficients of DCT; the wrinkle energy is calculated for different ages and is given an energy band to each age group. Unlike the previous method, which examined several regions, here the analysis is done primarily on the cheeks. The differences in energy of various age groups tends to be narrow hence log energies were considered for experimental results. The results showed that the proposed method was simple and easy to implement, was able to categorize the ages into various groups such as child, teenager and more generic labels such as young, mature and old.

A. Deepa and associate formulate a method wherein the estimation of the age is based on taking both the geometry-based quantitative analysis and texture. [2] The major steps taken include normalization, face cropping, filtering, feature extraction, classification. Due to the normalization step (done through histogram equalization) errors due to variation in illumination, and contrast deficiencies are resolved. Face cropping, with the help of Viola Jones- cascade object detector, which has the intrinsic advantage of a better detection rate. [15]The median filter, used for filtering, manages to retain the edge details and eliminate noise (even for images afflicted with faulty pixels; the intensity value of a "noisy" pixel does not affect the median value), without the drawback of increasing blur, which is the case with a lot of linear filters. Feature extraction, the most critical stage, is achieved by fragmenting the image, by means of the Two-Threshold Binary Decomposition (TBD) algorithm, into a group of binary images, whose texture pattern is obtained by deriving the fractal dimension of each group. SFTA (Segmentation-Based Fractal Texture Analysis) extracts texture features (represented in vector form) from the gray scale image (a gray scale image is used for better accuracy). SFTA has higher precision and accuracy for content based image retrieval (CBIR) and image classification; it is 3.7 times faster than Gabor and 1.6 times faster than Haralick methods.[3] The structure of the texture and its intensity transitions is represented by extracting the Local Directional Number Pattern (LDN), a 6-bit binary code assigned to each pixel of an input image, from facial regions. The face descriptor is a feature vector comprised of the extracted information. These features are concatenated into a feature vector, which is used as face descriptor. SFTA extracts features from gray-scale image. Classification is the last step and is done using a Deep Neural Network (DNN) to obtain the age.

Jana and Basu calculate wrinkle energy just like Ali, by counting white pixel in the Canny detector output of the ROI (namely forehead region, left /right eyelid region, middle of the eyebrows). [17] However, their method creates ratios of the number of white pixels (i.e. wrinkle densities) in different regions to create a value 'F', which is then sorted into an age group using Fuzzy C-Means algorithm. Each individual F value's membership in the different age clusters is calculated and the age cluster giving the highest membership value is assigned.

Imed Bouchrika et al have designed a method dependent on vision for age estimation.[6]

A combination of features comprised of local and global characteristics is formulated using LBP (Local Binary Pattern). From the chain of locally formed histogram vectors from grid cells of face images, a feature histogram is made. Classification follows a tress-based pattern. The feature vector for age estimation is obtained with the help of detecting the face via the Viola and Jones algorithm, like in Deepa's method. Further refinement is done by removing unimportant features such as hair, which may help counter the limitation of the former. Rather than subjecting all the age groups to a one time feature selection, feature selection is applied recursively at each level to generate subset of features.

Louis Quinn and Margaret Lech have pioneered a multi-stage binary age estimation (MSAE) system that is built upon a structure of Neural Network (NN) and Support Vector Machine (SVM).[16] In the image feature extraction step, the orthogonal locality projection (OLPP) and Sobel Edge Detector (SED) features are derived. OLPP was first designed as an facial recognition method based on Fischer faces or locality preserved projections (LPP). SED is used to identify wrinkles, lines etc which acts as indicators of aging. Sobel operator is used for edge detection. Due to the convoluted nature of the age recognition task the newly classified face may not remain in the allocated sub-space (age-group) defined by LDA and OLPP, therefore the MSAE system must adapt such that the overall accuracy remains the same.

At each step of the classification procedure, a binary decision is used to test if the face is similar to any one of the overlapping age groups. In the beginning the gap between the age–groups will be broad, but over time as the algorithm progresses the gap becomes narrower, until convergence, providing the final age group. The MSAE process uses four classifiers in order to reinforce the decision process. A pair of ANN's and SVM units each is used. The training is based on gradient of steepest descent algorithm, created and tested using the MORPH2 database [14].

Shin, Seo and Kwon have devised an age and gender estimation system that considers feature-based distinctions in face images using a Convolution Neural Network (CNN) and Support Vector Machine (SVM).[18] Due to the use of ethnicity based databases for training, the system shows limited performance. Initially in the system's methodology, the ethnicity of the facial image is determined by a CNN trained with manually collected face images. Then, the determined ethnicity is used to select an SVM classifier for the final age and gender estimation. Novelty of the experiment includes the success in generalizing the CNN for the classification of the ethnicity and the improved performance in age and gender prediction compared to other methods. Also, the gender classification accuracy was not affected by the method, even though there were differences in the ethnicity of facial images.

Lee and associates advocate a deep residual learning model for the estimation of age and gender. [12] The method estimates age and gender of each faces by detecting faces that are fed as input images. The proposed model is trained with the images in IMDB-WIKI database. First, the baseline model is designed for performing comparison with the other models. The baseline model is a simple regression model and it gives estimated age and gender simultaneously. Data augmentation is used to train the model and it showed better performance when compared to the model without data augmentation. Also, a regression model is adopted here rather than using a classification model as it yields an improved performance. Investigations from this method showed that real age and gender estimation is much effective than using any other estimation methods and the residual connection help to improve its performance.

Hosseini presents a convolution neural network (CNN) based architecture for joint age-gender classification. [9] Back propagation is used for the learning process of Gabor filter responses. The architecture is trained to label the input images into eight 1 ranges of age and two types of gender. Adience Dataset is used here for training. The system showed improvement up to 7 percent in age accuracy and 2 percent in gender accuracy when compared to other estimation methods. The advantage of the scheme is to let the deep neural network to focus on useful features, which ultimately improved the performance of the system.

Koichi et al have designed an age and gender prediction method from face images using CNN.[10] This method makes use of public face databases for its experiment. The deep multi task learning in the method showed an accuracy of about 93.54% in case of gender prediction. However, A powerful data cleaning method has not been used here, which can improve the accuracy of the method.

After obtaining the age and gender of a person, our recommendation system can determine which demographic group they belong to, and then display the preset recommendations which is targeting that group.

6. RESPONSE MEASUREMENT

Another vital aspect of online recommendation based systems is the response aspect. Based on click rate, page views and purchases, corporations are able to determine how accurately the system is predicting user behavior, understand market trends and also gauge how effective their own marketing strategies are.

To replicate this facility in the proposed system, emotion tracking can be utilized. The customer's emotion will be identified after viewing the suggested content, classified (for example, categorized into a "negative", "positive" or "neutral" emotion). This will serve as a rudimentary litmus test to see if the consumer considers the recommendation to be apt and useful. Of course, the extent of how useful this mode of response measurement turns out to be is dependent on the accuracy of the emotion detection method used.

A Google-developed TensorFlow library is utilized by Lee and Hong to modify a CNN to conform to the respective database, and accomplished an emotion recognition system (with the aid of deep learning). [8] There were two modes of recognition: a binary label mode (classifying an emotion as positive or negative) and a seven label mode (classifying an emotion as surprised, happy, angry, sad, scared, disgusted, or neutral). An Arousal Valance model is used to classify each test image based on their generated value in the CNN. Datasets were created by clipping images from video (containing expressions corresponding to the seven emotions) .

Vithanawasam and Madhusanka focused on creating face (and upper body) emotion recognition specifically for service robots. [19] They used Fischerfaces, an algorithm for face recognition based on the LDA (Latent Dirichlet allocation) .Fisherfaces, which dichotomizes the Eigenspace into classes, has the benefit that once the Eigen Space has been defined (or learned), the recognition can run in real-time, which is a prerequisite in the case discussed in this paper. It gives better classification results, which offsets its complexity and greater processing time. It remains to be seen if the greater processing time will greatly delay the system overall. The proposed recognition classifier is able to differentiate anger, fear, neutral and bored emotions.

After obtaining the emotion, the system can take record the observation. Later, the system can do a demographic based statistical analysis to show how each group responded to each recommendation (basically, what was the percentage positive/negative response, were there any content with a high positive response from a non-target group etc.) Analyzing this data, along with purchase data in the case of a retail scenario can help managerial authorities in course correction, and help the system designers understand the predictive accuracy of the system.

7. CONCLUSION

To conclude, it is possible to create a physical version of an online recommendation system to invigorate retail and entertainment spaces by providing personalized content by using image processing techniques. Just by examining a person's face (the most immediate and available source of data) the system can understand a lot about them and make a very likely guess as to their interest. As mentioned, the concerns for the techniques are data security, accuracy of recognition and speed of recognition. The system can be further advanced by providing facial identification (i.e. identifying exactly who the customer is). This will allow the system to track past history of the customer, social media activity, and thus help make a much more informed prediction.

REFERENCES

[1] A. A. Shejul, K. S. Kinage and B. E. Reddy, "Comprehensive review on facial based human age estimation," *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, Chennai, 2017, pp. 3211-3216.doi:10.1109/ICECDS.2017.8390049

[2]A.Deepa, and T.Sasipraba. Age Estimation in Facial Images Using Histogram Equalization.IEEE,19June2017,ieeexplore.ieee.org/document/7951767.

[3] A. F. Costa, G. Humpire-Mamani and A. J. M. Traina, "An Efficient Algorithm for Fractal Analysis of Textures," 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images, Ouro Preto, 2012, pp. 39-46. doi: 10.1109/SIBGRAPI.2012.15

[4] Ali, Syed Musa, et al. Age Estimation from Facial Images Using Biometric Ratios and Wrinkle Analysis. IEEE, 6 Aug. 2015, ieeexplore.ieee.org/document/7176403.

[5] Bekhouche, Salah Eddine, et al. "Facial Age Estimation and Gender Classification Using Multi Level Local Phase Quantization." 2015 3rd International Conference on Control, Engineering & Information Technology (CEIT), 2015, doi:10.1109/ceit.2015.7233141.

[6] Bouchrika, Imed, et al. Automated Clustering and Estimation of Age Groups from Face Images Using the Local Binary Pattern Operator.IEEE, 25 Feb.2016,ieeexplore.ieee.org/document/7416714.

[7] Deng, Houtao. "Recommender Systems in Practice." Towards Data Science, Medium, 13 Feb.2019,towardsdatascience.com/recommender-systems-in-practice-cef9033bb23a.

[8] H. Lee and K. Hong, "A study on emotion recognition method and its application using face image," *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju, 2017, pp. 370-372.URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8191005&isnumber=8190740

[9] Hosseini, Sepidehsadat, et al. "Age and Gender Classification Using Wide Convolutional Neural Network and Gabor Filter." 2018 International Workshop on Advanced Image Technology(IWAIT),2018,doi:10.1109/iwait.2018.8369721.

[10] Ito, Koichi, et al. "Age and Gender Prediction from Face Images Using Convolutional Neural Network." 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference(APSIPAASC),2018,doi:10.23919/apsipa.2018.8659655.

[11] Khan, Sahib, et al. Wrinkles Energy Based Age Estimation Using Discrete Cosine Transform. IEEE, 21Jan.2016,ieeexplore.ieee.org/document/7389184.

[12] Lee, Seok Hee, et al. "Age and Gender Estimation Using Deep Residual Learning Network." 2018 International Workshop on Advanced Image Technology (IWAIT), 2018, doi:10.1109/iwait.2018.8369763.

[13] Mishra, Madhav. "Learning How Recommendation System Recommends." Towards Data Science, Medium, 9 Aug. 2018, towardsdatascience.com/learning-how-recommendation-system-recommends-45ad8a941a5a

[14] MORPH Non-Commercial Release Whitepaper. [Online] Available:

http://www.faceaginggroup.com

[15] P. Viola and M. Jones, 'Rapid Object Detection using a Boosted Cascade of Simple Features', in Conference on Computer Vision and Pattern Recognition, 2001.

[16] Quinn, Louis, and Margaret Lech. Multi-Stage Classification Network for Automatic Age Estimation from Facial Images. IEEE, 28 Jan. 2016, ieeexplore.ieee.org/document/7391782/.

[17] R. Jana and A. Basu, "Automatic age estimation from face image," *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Bangalore, 2017,pp.87-90.

[18] Shin, Minchul, et al. "Face Image-Based Age and Gender Estimation with Consideration of Ethnic Difference." 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2017, doi:10.1109/roman.2017.8172359.

[19] Vithanawasam, Tharinda & Madhusanka, Achintha. (2018). Dynamic Face and Upper-Body Emotion Recognition for Service Robots. 10.1109/ICIS.2018.8466505.