Transgender Face Recognition using ROI based Convolutional Neural Network

R. Bhuvaneswari
Assistant Professor
Department of Computer Science and Engineering
Sri Venkateswara College of Engineering

S. Ganesh Vaidyanathan
Professor
Department of Electronics and Communication Engineering
Sri Venkateswara College of Engineering

ABSTRACT

This framework deals with a distinct problem domain where face recognition is performed for individuals who have undergone gender transformation over a period of time. The recognition rate of face recognition stands a major challenge in dealing with the pictures or video frames of transgender individuals. Typically the sexual orientation change causes serious modifications in the actual appearance of the face just as in the body of a transgender individual. Therefore, it presents extra complexity / burden in taking care of the accuracy as far as transsexual face acknowledgement. Subsequently, there is a requirement for face recognition framework to reliably distinguish the people after they go through sex change. As Convolutional Neural Network (CNN) has demonstrated to be one of the powerful tool in dealing with feature extraction in images, a new framework is presented which uses CNN to increase the recognition rate in transgender images. The proposed model extracts the features of transgender's face components such as two eyes, nose and mouth using CNN. The CNN have been utilized in the proposed model. The investigations were done on HRT transsexual database.

General Terms

Face Recognition, Region of Interest, Feature Extraction

Keywords

Transgender, Convolutional Neural Network, Support Vector Machine

1. INTRODUCTION

High security is needed for the data and for the information that has been collected and maintained by us. Biometrics has now gotten more consideration among the existing security methods. Face Biometrics has been proven as one of the best biometric methods as this has been handled with the help of the images without touching or interacting with the person. Even the accuracy obtained in face recognition found to be good under some restricted condition, it still finds a challenging task under some unconstrained condition. The challenges includes variation due to make-up [1], aging [8], Plastic surgery [10], change in skin texture [11] [6]. In addition to these parameters which hurdles the face recognition accuracy, the physical changes in the face during the gender transition results in additional challenge to the existing face recognition techniques. The gender transition is supported through the medical procedure known as Hormone Replacement Therapy(HRT). Using this therapy, the hormones of opposite sex is induced thereby suppressing the present hormone in a person's body. Due to this, there can be a intensive change in the facial features. For those persons, during the gender transformation through HRT have major variation in the facial shape and texture [7] [3] [4]. One of the challenges in face recognition that has been in recent times is facial changes due to the gender transformation over a period of time through hormone replacement therapy (HRT) which acts face fat distribution thus resulting in changes to face shape and texture. The texture variation [9] results as the skin get thinned when the gender transformation is from male to female and the skin get thickened when the gender transformation is from female to male. The shape changes due to HRT are caused due to the changes in the fat distribution. The problem addressed in the literature studies in [2] [5] have been limited for transgender face recognition. The objective of this work is to extract the features that has to give high descriptor ability as well as less in dimension for face recognition across gender transformation in surveillance systems and biometric obfuscation. Also, the proposed model handles the variation in facial shape, size, features in transgender face recognition. Gender transformation results in variations in face shape and texture which result in great challenge in face recognition. Hence a better feature representation is needed to address the issues in recognizing facial features. To handle the problem of face recognition under gender transformation, a face component model using CNN has been proposed in this paper to improve the accuracy in transgender face recognition.

2. BACKGROUND

Convolutional Neural Network (CNN) is a feed-forward neural network which extracts the features using many convolution layer. A CNN consists of (a) Convolution layers (with the set of kernels/ filters), (b) Pooling layer and (c) Output layer. The input of CNN is of original input data without any pre-processing techniques. The number of nodes in the input layer is determined by the dimension of the input data. The input is fed into the convolution layer. The computation in the convolutional layer with the kernel/filters and bias extracts the most discriminate features from the input which is

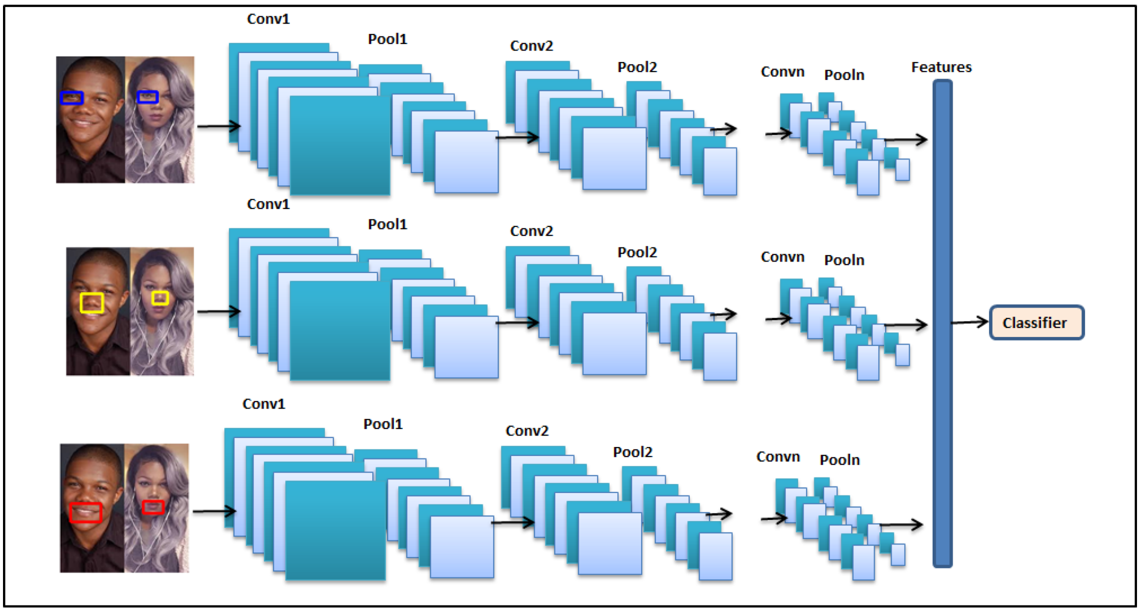

Fig. 1. Proposed ROI based Convolutional Neural Network.

described by

$$X = Conv(I, F) + bias \tag{1}$$

where Conv(.) represents the convolution function, I is the input image, F is the kernel/filter applied and X is the convolved image. The generated feature maps are applied to a nonlinear activation functions, such as rectified linear unit (ReLu),sigmoid function, tanh function

$$M = AF(X) \tag{2}$$

where AF (.) represents the activation function and M is the feature map. The various activation functions are as follows: The rectified linear unit (ReLu) is the most commonly used activation function in many medical applications, especially in CNN. The ReLu activation function is given by

$$R(x) = max(0, x) \tag{3}$$

The Sigmoid activation function is given by

$$\sigma(x) = \frac{1}{(1 + e^x)} \tag{4}$$

The tanh activation function is given by

$$tanh(x) = \frac{2}{(1 + e^{2x}) - 1} \tag{5}$$

The convolved image is then passed to the pooling layer. The pooling operation aims with (i) Reduction in dimensionality (ii) Avoid over-fitting of data (iii) Reducing the number of parameters thereby reducing the computational cost. In the pooling layer, either maximum pooling or minimum pooling or an average pooling can be employed. After the pooling layer, fully connected layer is added. The model will get trained with the features in the fully connected layer by a classifier. Then the output layer provides the results of classification.

3. REGION OF INTEREST (ROI) BASED CONVOLUTIONAL NEURAL NETWORKS FOR TRANSGENDER FACE RECOGNITION

This section describes the proposed model for the face recognition of transgender images. There are many challenges associated with extracting information from face that includes aging, pose, illumination and expression which is described as A-PIE and even more challenging in recent times is identifying or recognizing a person's face that has undergone gender transformation over a period of time through hormone replacement therapy. Hence, the proposed work first deals with the selection of region of interest (ROI) in the transgender face in order to get the best descriptor ability.

Some studies have been carried out about the transformation regions in the face among the transgender over the gender transformation period. As there can be significant less transformation in the regions of eyes, nose and mouth when compared to the other regions in the transgender face, the proposed work aims to extract the features of two eyes, nose and mouth. Figure 1 shows the architectural details of the proposed ROI based Convolutional Neural Network.

Let the training images contains both the images of before and after gender transformation. Let the dimension of detected eyes, nose and mouth from all the 'n' training images be $\{LE_{x_1} \times LE_{y_1}, RE_{x_1} \times RE_{y_1}, ..., LE_{x_n} \times LE_{y_n}, RE_{x_n} \times RE_{y_n}\}$, $\{N_{x_1} \times N_{y_1}, ..., N_{x_n} \times N_{y_n}\}$ and $\{M_{x_1} \times M_{y_1}, ..., M_{x_n} \times M_{y_n}\}$ respectively.

$$(E_{w_1} \times E_{h_1}), (E_{w_2} \times E_{h_2}) = \left(\frac{\sum_{i=0}^{n} LE_{x_i} \times LE_{y_i}}{n}\right), \left(\frac{\sum_{i=0}^{n} RE_{x_i} \times RE_{y_i}}{n}\right) \tag{6}$$

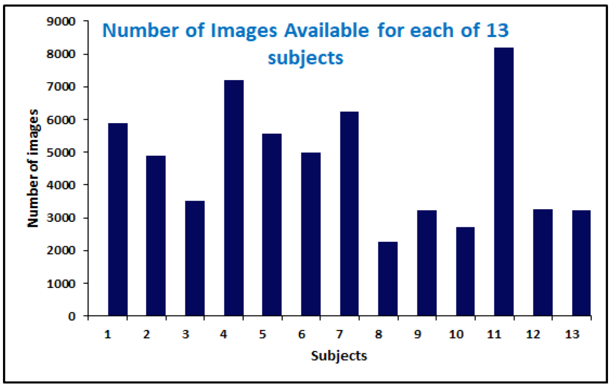

Fig. 2. Number of Images Available for each of 13 subjects.

$$(N_w \times N_h) = \frac{\sum_{i=0}^{n} N_{x_i} \times N_{y_i}}{n} \quad (7)$$

$$(M_w \times M_h) = \frac{\sum_{i=0}^{n} M_{x_i} \times M_{y_i}}{n} \quad (8)$$

Let the resized dimension of detected eyes, nose and mouth be given through the equation 6 to 7. Hence,$(E_{w_1} \times E_{h_1}), (E_{w_2} \times E_{h_2}), (N_w \times N_h)$ and $(M_w \times M_h)$ is the resized dimension of left eye, right eye, nose and mouth respectively. For each region of interest, a dedicated CNN is built which aims at extracting the high descriptor features.
Let the set of training samples be denoted as T= (X_n, y_t) where X_n are the set of feature vectors generated for each image and y_t is the corresponding class label. Training is done using support vector machine(SVM).
Support vector machine (SVM) is a binary classifier. Let L_s be the total number of support vectors and $\{\alpha_j^*\}_{j=1}^{L_s}$ be the optimal Lagrangian coefficients obtained after training the SVM. The discriminant function of SVM for an input example $\boldsymbol{X}$ is given by

$$D_{SVM}(\boldsymbol{X}) = \sum_{j=1}^{L_s} \alpha_j^* y_j K(\boldsymbol{X}, \boldsymbol{X}_j) \quad (9)$$

The class label of $\boldsymbol{X}$ is determined using the sign of the value of $D_{SVM}(\boldsymbol{X})$.

4. DATASET

The experiment has been done using the Hormone Replacement Therapy (HRT) dataset [5]. The proposed work has been carried out with 29475 images which corresponds to the images of 13 persons. Figure 2 shows the number of images available for each of these 13 subjects. In this, 28575 images and 900 images have been used for training and testing images respectively. Table 2 shows the number of images available for each subject. For the subject 11, nearly 8000 images were available whereas for the subject 8 less than 3000 images are available. These images are the collection of these subjects throughout the transformation stages.

5. EXPERIMENTS AND RESULTS

Once the region of interest have been detected from all the images, using Eq. 6 to 8 the dimension of these regions have been resized or fixed whose values have been given in Table 1. Each dedicated CNN is used in the proposed model which handles each ROI (Eyes, Nose and Mouth). These ROI based CNN have been utilised in feature extraction in the proposed model. The parameters of all the CNN for each ROI have been depicted from the Table 2 to Table 4. As the CNN has the high descriptive feature extraction ability, the proposed model utilises the CNN as the feature extraction technique. The extracted features from all ROI based CNN have been used for training the model which uses the SVM classifier for recognition task.

Table 1. Dimension of ROI

ROI	Dimension
Eyes $(E_{w_1} \times E_{h_1}), (E_{w_2} \times E_{h_2})$	38×23
Nose $(N_w \times N_h)$	38×32
Mouth $(M_w \times M_h)$	50×22

Table 2. Dimension of feature maps corresponds to the CNN which deals with ROI: Eyes

Type	Maps
Input Layer	38×23
Conv1 Layer	28 maps of 38×23 neurons
Pool1 Layer	28 maps of 19×12 neurons
Conv2 Layer	28 maps of 19×12 neurons
Pool2 Layer	28 maps of 10×6 neurons
Conv3 Layer	28 maps of 10×6 neurons
Pool3 Layer	28 maps of 5×3 neurons

Table 3. Dimension of feature maps corresponds to the CNN which deals with ROI: Nose

Type	Maps
Input Layer	38×32
Conv1 Layer	28 maps of 38×32 neurons
Pool1 Layer	28 maps of 19×16 neurons
Conv2 Layer	28 maps of 19×16 neurons
Pool2 Layer	28 maps of 10×8 neurons
Conv3 Layer	28 maps of 10×8 neurons
Pool3 Layer	28 maps of 5×4 neurons

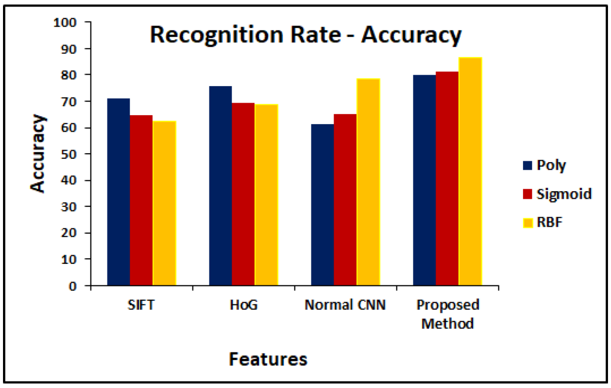

Fig. 3. Comparison of Accuracy of Recognition rate.

Table 4. Dimension of feature maps corresponds to the CNN which deals with ROI: Mouth

Type	Maps
Input Layer	50×22
Conv1	28 maps of 50×22 neurons
Pool1	28 maps of 25×11 neurons
Conv2	28 maps of 25×11 neurons
Pool2	28 maps of 13×6 neurons
Conv3	28 maps of 13×6 neurons
Pool3	28 maps of 7×3 neurons

Table 5. Comparison of Accuracy obtained using Proposed model with different features and different kernels in SVM.

	Accuracy rate Obtained using			
Kernel	SIFT Features	HoG Features	Normal CNN	Proposed Model
Poly	71.02%	75.57%	61.22%	**80.01**%
Sigmoid	64.83%	69.56%	65.23%	**81.25**%
RBF	62.21%	68.86%	78.66%	**86.52**%

The accuracy rates have been compared whose results are shown in the Table 5. The proposed model have been analysed with the various kernel in the SVM classifier. The best results have been obtained with the kernel 'RBF'. The recognition rate in the proposed model found to be good with the accuracy rate of 86.52% when compared to the other models. Fig. 3 shows the diagrammatic representation of the accuracy rate of the proposed model with the various kernels. The features obtained using SIFT, HoG and normal CNN also have been trained using the SVM classifier, but the accuracy rate obtained is not good enough when compared to the proposed model.

6. CONCLUSION

Transgender face recognition is the emerging challenge in the field of face biometric. During the gender transformation process, the subtle changes to the skull bone results challenging with the identification process of transgender persons. In the proposed model the face components such as eyes, nose and mouth have been utilized as these features have less impact in the physical change during the transformation period. As CNN finds to be a best tool in many image classification / recognition tasks, it has been incorporated in the proposed model to achieve the fine accuracy. The future work aims to work with suitable deep learning algorithm which best suits for improving the accuracy rate still further.

7. REFERENCES

[1] Antitza Dantcheva, Cunjian Chen, and Arun Ross. Can facial cosmetics affect the matching accuracy of face recognition systems? In *2012 IEEE Fifth international conference on biometrics: theory, applications and systems (BTAS)*, pages 391–398. IEEE, 2012.

[2] Vijay Kumar, Ramachandra Raghavendra, Anoop Namboodiri, and Christoph Busch. Robust transgender face recognition: Approach based on appearance and therapy factors. In *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pages 1–7. IEEE, 2016.

[3] Haibin Ling, Stefano Soatto, Narayanan Ramanathan, and David W Jacobs. Face verification across age progression using discriminative methods. *IEEE Transactions on Information Forensics and security*, 5(1):82–91, 2009.

[4] Gayathri Mahalingam and Chandra Kambhamettu. Face verification of age separated images under the influence of internal and external factors. *Image and Vision Computing*, 30(12):1052–1061, 2012.

[5] Gayathri Mahalingam, Karl Ricanek, and A Midori Albert. Investigating the periocular-based face recognition across gender transformation. *IEEE Transactions on Information Forensics and Security*, 9(12):2180–2192, 2014.

[6] Raghavendra Ramachandra, Kiran Raja, Sushma Venkatesh, and Christoph Busch. Collaborative representation of statistically independent filters' response: An application to face recognition under illicit drug abuse alterations. In *Scandinavian Conference on Image Analysis*, pages 448–458. Springer, 2017.

[7] Narayanan Ramanathan and Rama Chellappa. Face verification across age progression. *IEEE transactions on image processing*, 15(11):3349–3361, 2006.

[8] Karl Ricanek and Tamirat Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pages 341–345. IEEE, 2006.

[9] P-G Sator, JB Schmidt, MO Sator, JC Huber, and H Hönigsmann. The influence of hormone replacement therapy on skin ageing: a pilot study. *Maturitas*, 39(1):43–55, 2001.

[10] Richa Singh, Mayank Vatsa, Himanshu S Bhatt, Samarth Bharadwaj, Afzel Noore, and Shahin S Nooreyezdan. Plastic surgery: A new dimension to face recognition. *IEEE Transactions on Information Forensics and Security*, 5(3):441–448, 2010.

[11] Daksha Yadav, Naman Kohli, Prateekshit Pandey, Richa Singh, Mayank Vatsa, and Afzel Noore. Effect of illicit drug abuse on face recognition. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–7. IEEE, 2016.