# Robust Transgender Face Recognition: Approach based on Appearance and Therapy Factors

Vijay Kumar † R. Raghavendra * Anoop Namboodiri † Christoph Busch*

*Norwegian Biometrics Laboratory,
Norwegian University of Science and Technology (NTNU), 2802 Gjøvik, Norway

†International Institute of Information Technology (IIIT) Hyderabad, India.
{raghavendra.ramachandra, christoph.busch}@ntnu.no {(vijaykumar.r@research, anoop@} iiit.ac.in

## Abstract

*Transgender face recognition is gaining increasing attention in the face recognition community because of its potential in real life applications. Despite extensive progress in traditional face recognition domain, it is very challenging to recognize faces under transgender setting. The gender transformation results in significant face variations, both in shape and texture gradually over time. This introduces additional complexities to existing face recognition algorithms to achieve a reliable performance. In this paper, we present a novel framework that incorporates appearance factor and a transformation factor caused due to Hormone Replacement Therapy (*HRT*) for recognition. To this extent, we employ the Hidden Factor Analysis (*HFA*) to jointly model a face under therapy as a linear combination of appearance and transformation factors. This is based on the intuition that the appearance factor captures the features that are unaffected by the therapy and transformation factor captures the feature changes due to therapy. Extensive experiments carried out on publicly available* HRT *transgender face database shows the efficacy of the proposed scheme with a recognition accuracy of* 82.36%.

## I. Introduction

Face recognition has been intensively studied in the field of Biometrics for more than five decades. In spite of these extensive efforts from both academic and industrial research to improve the reliability and robustness, it still remains a challenging problem because of high variability faces exhibit in terms of pose, illumination, age, etc. In addition to those widely discussed challenges, faces under plastic surgery [9], disguise [1], 3D masks for spoofing [7] [6] and gender transformation [5] present even more challenging conditions which limits the robustness of the existing face recognition systems. Among the various emerging challenges on face recognition systems, the most recent and unique challenge is the recognition under gender transformation (or Transgender face recognition).

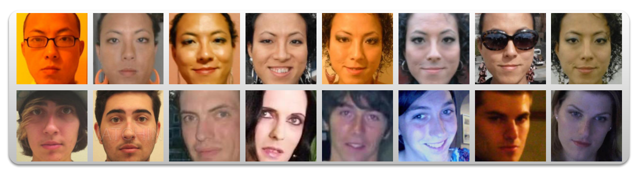


Fig. 1: Images from HRT database. Row 1 shows images of one and the same subject at various time periods after the gender transformation. Apart from regular variations seen in faces such as illumination, expression, occlusion, pose, there is an extreme change in the facial characteristic due to gender transformation. Similarly, row 2 shows the effect of transformation on other 4 subjects from the database.

Gender transformation is normally carried out through Hormone Replacement Therapy (HRT) that replaces the natural sex hormone of the subject with its opposite sex. Thus, the gender transformation will affect the face distribution due to changes in its overall shape and texture [5]. Depending upon the type of the gender transformation (i.e. from male to female or vice versa), the most significant changes in the transformed face can be noticed in terms of fine wrinkles and lines, stretches in the dermis, skin thickening or thinning and texture variations. All these changes increase the intra-class variations significantly leading to

Vijay Kumar and R. Raghavendra have contributed equally to this work.

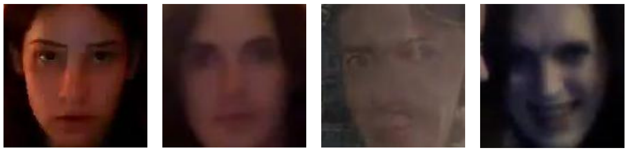


Fig. 2: Illustration of bad quality images from the HRT database collected from YouTube videos.

performance degradation of conventional face recognition systems. We show a few examples from the HRT database in Figure 1 to illustrates these challenges. First row shows images of one and the same subject captured in a period of $36$ months after the therapy. Similarly, the second row in Figure 1 shows image pairs of $4$ different subjects before and after HRT that further demonstrates the complexity of the problem for robust face recognition. It is clear that, the therapy introduces a very high degree of intra-class variation making it an unique problem compared with other challenges like plastic surgery [9] or face disguise [1].

Even though face recognition systems have been applied for a wide spectrum of real life challenges, the gender transformation (or transgender) problem is an upcoming problem that introduces new challenges for face recognition. The main challenge in the transgender face recognition is the change in shape/texture features of the subject's face that happens significantly during the transformation period. Since these changes occur over time, it is very challenging to achieve a consistent face recognition as the aging component also plays a vital role in the recognition accuracy.

There are very few studies in the literature focused on the Transgender face recognition [4], [5]. Authors in [4] conduct a preliminary study on transgender face recognition by exploring the eye (or periocular) region since the periocular region is less effected compared to other facial regions. They use $4$ different features namely TP-LBP, LBP, HoG and SIFT to achieve a reasonable recognition performance. The experiments conducted on the Transgender face database with $11$ subjects demonstrated the robustness based on the periocular region compared to the entire face with an average improvement of over $4\%$. The work is extended in [5], by including both nose and mouth region in addition to two periocular regions which further improves the performance. However, both these approaches consider only appearance features, in four different parts (mouth, nose and two periocular) while ignoring the changes in face appearance over time. Furthermore, the use of only appearance based features will demand the need for a high quality face images where periocular, mouth and nose region are clearly visible. This is quite difficult to obtain in real-life scenarios, few examples of which are shown in Figure 2.

In this work, we propose a novel solution for transgender face recognition by exploring both appearance factor, and transformation factor caused due to Hormone Replacement Therapy (HRT). Unlike the previous schemes, the proposed framework is based on the entire face region obtained in highly unconstrained scenarios along with the time information to achieve the robust transgender recognition. Our approach is based on the observation that the therapy introduces a significant change in the facial characteristic over time. Thus if we model such therapy induced changes over time, better performance can be achieved by separating therapy component from the appearance or identity component that are invariant over time. To this extent, our proposed scheme will explore the Hidden Factor Analysis (HFA) [2] which is a generative model to represent the face as a linear combination of appearance and therapy components. The basis for these components are learned through an Expectation-Maximization (EM) algorithm from a training set of images containing both time label that indicates the time period elapsed since the therapy and an identity label. To summarise, we make the following main contributions in the paper:

- A novel framework for transgender face recognition by exploring the Hidden Factor Analysis (HFA) by considering both appearance and therapy components. To our knowledge, this method is explored for the first time on the transgender face recognition.
- Extensive experiments are carried out on publicly available HRT database [4], [5] and collected data from YouTube videos showing significant improvement over current state-of-the-art face recognition algorithms namely, Sparse Representation Classifier (SRC) [12] and Collaborative Representation Classifier (CRC) [13].

The rest of the paper is organized as follows: Section II describes the Hidden Factor Analysis tailored for the transgender face recognition, Section III presents the proposed framework for transgender face recognition, Section IV presents the experimental results and discussion. Finally, Section V draws the conclusion.

## II. Hidden Factor Analysis of transgender face recognition

In this section, we present the Hidden Factor Analysis (HFA) [2] and show how it can be adapted for transgender face recognition application. We consider the problem of face recognition that is invariant to gender transformation through a hormonal transformation therapy (HRT). As shown in the Figure 1, this is an extremely challenging problem since the face of a subject under consideration undergoes rapid transformation due to the therapy. Previous approaches proposed in [3], [5] do not account for such transformations occurring in faces due to HRT.

In this paper, we propose an approach that models the face of a subject under HRT into two factors - an appearance factor and transformation or therapy factor caused by the therapy. This is based on the intuition that the faces belonging to the same subject but were captured at different periods after the therapy may contain common appearance features that are stable and consistent even after the therapy. Also, the type of transformation seen in different subjects at different time periods after the therapy are usually similar. For example. one could see a change in skin texture in the majority of the subjects after the first few months after therapy. If one could learn such transformations due to therapy at different time periods after therapy, we can separate out the face into an stable and discriminative appearance component and a noisy therapy component, and exploit them for superior recognition.

Our approach is highly inspired from the recently proposed algorithm for age-invariant face recognition based on hidden factor analysis [2]. One of our major contributions is to demonstrate the successful application of the state-of-the-art age-invariant face recognition algorithm for the transgender face recognition. Following [2], we model the face with an appearance and transformation component due to therapy using hidden factor analysis (HFA) and capture these components in two independent hidden or latent factors, namely *appearance* and *therapy* factors. Given a face $x \in \mathbb{R}^d$ under therapy, we consider a linear generative model and represent it as a linear combination of appearance and therapy components as follows,

$$x = \mu + A\alpha + B\beta + \eta \tag{1}$$

where $\alpha \in \mathbb{R}^p$ represents the latent appearance factor and $\beta \in \mathbb{R}^q$ is the latent transformation factor due to therapy. $\mu \in \mathbb{R}^d$ is the mean feature over all the samples and $\eta \in \mathbb{R}^d$ is a additive Gaussian noise with zero mean and variance $\sigma^2$. $A \in \mathbb{R}^{d\times p}$ represents a basis that spans the appearance of various subjects while $B \in \mathbb{R}^{d\times q}$ captures the transformation changes due to therapy.

The parameters $\theta = (A, B, \mu, \sigma^2)$ are estimated from the training data through maximum likelihood estimation using an Expectation-Maximization (EM ) algorithm. We present below the final results for the parameters that are updated in E-step and M-step, respectively. We refer the interested readers to the original paper [2] for detailed proof of the algorithm. Let $N_i$ and $M_k$ be the number of samples belonging to $i$-th subject and $k$-th period, respectively. Let $N$ denote the total number of samples.

Given the training set $X = \{x_i^k, i = 1, \ldots, N_i, k = 1, \ldots, M_k\}$ where $x_i^k$ denote a sample belonging to $i$-th subject at $k$-th time period group after the gender transformation therapy. The mean sample $\mu$ is estimated from the training set as

$$\mu = \frac{1}{N}\sum_{i,k} x_i^k \tag{2}$$

**E-step:** Given initial parameters $\theta_0$, and a set of training data $X$, the sufficient statistics are estimated.

$$\alpha_i = \frac{A^T\Sigma^{-1}}{N_i}\sum_{k=1}^{N_i}(x_i^k - \mu) \tag{3}$$

$$\beta_k = \frac{B^T\Sigma^{-1}}{M_k}\sum_{i=1}^{M_k}(x_i^k - \mu) \tag{4}$$

$$E_{x_i x_i^T} = \frac{(I - A^T\Sigma^{-1}A)}{N_i} + x_i x_i^T \tag{5}$$

$$E_{y_k y_k^T} = \frac{(I - B^T\Sigma^{-1}B)}{M_k} + y_k y_k^T \tag{6}$$

$$E_{y_k x_i^T} = \frac{(-B^T\Sigma^{-1}A)}{\sqrt{N_i M_k}} + y_k x_i^T \tag{7}$$

$$E_{x_i y_k^T} = \frac{(-A^T\Sigma^{-1}B)}{\sqrt{N_i M_k}} + x_i y_k^T \tag{8}$$

where

$$\Sigma = \sigma^2 I + AA^T + BB^T$$

**M-step:** Given the initial parameters $\theta_0$, the parameters $\theta$ that maximizes the log-likelihood function is estimated using the sufficient statistics estimated in E-step,

$$M_a = \sum_{i,k} E_{x_i x_i^T} \tag{9}$$

$$M_b = \sum_{i,k} E_{y_k x_k^T} \tag{10}$$

$$M_c = \sum_{i,k}(x_i^k - \mu)\alpha_i^T \tag{11}$$

$$M_d = \sum_{i,k}(x_i^k - \mu)\beta_k^T \tag{12}$$

$$M_e = \sum_{i,k} E_{y_k x_i^T} \tag{13}$$

$$M_f = \sum_{i,k} E_{x_i y_k^T} \tag{14}$$

$$U = (M_c - M_d M_b^{-1} M_e)(M_a - M_f M_b^{-1} M_e)^{-1} \tag{15}$$

$$V = (M_d - M_c M_a^{-1} M_f)(M_b - M_e M_a^{-1} M_f)^{-1} \tag{16}$$

$$\sigma^2 = \frac{1}{Nd}\sum_{i,k}[(x_i^k - A\alpha_i - B\beta_k - \mu)^T(x_i^k - \mu)] \tag{17}$$


Fig. 3: Block diagram of the proposed framework for transgender face recognition

## III.Proposed framework for Transgender Face Recognition

Figure 3 shows the block diagram of our proposed framework for the transgender face recognition containing the following components:

### A.Face Detection and Normalization

Given the video frame, we first perform the face detection and normalization. In this work, the face detection is carried out using the Viola-Jones algorithm [11] by considering its robustness and performance in a real-time scenario. Since most of the face images in HRT database are captured by the subjects on their own, it exhibits variation in face pose as well as challenging backgrounds and illumination. Thus, the use of Viola-Jones algorithms will result in numerous false detection that can be addressed as mentioned in [8]. Those false positives that cannot be mitigated using the technique described in [8] are then visually determined and manually rectified to improve the overall performance of the system.

### B.Feature Extraction

Once the face is detected, the images are re-sized to a fixed size of $120 \times 120$ and local features are extracted. We represent the images using 3 types of features namely, Local Binary Patterns (LBP), Histogram of Oriented Gradients (HOG) and dense-SIFT. LBP and HOG features are extracted using a cell-size of 12. For dense-SIFT, we extract the patches densely of size $96 \times 96$ (or equivalently bin size of 24 with $4 \times 4$ cells) and compute SIFT features from each of these patches. We used the publicly available software VLFEAT [10] for extracting these features. Once the features are extracted, we reduce the dimension of these features using Principal Component Analysis (PCA).

### C.Comparison using HFA model

After we perform the feature extraction and dimensionality reduction on samples, we consider the samples corresponding to the training set to learn the basis A and B using HFA model explained in the section II. Finally, the representation $y$ of any sample $x$ is obtained by projecting it on the appearance component $A$ alone as,

$$y = AA^T \Sigma^{-1}(x - \mu) \tag{18}$$

During testing, the target images are passed through the same pipeline and finally the matching is performed using the cosine distance [2].

## IV.Experiments and Results

In this section, we present the details of the database employed for experiments, performance evaluation protocol and discussion on the obtained results.

### A.Dataset

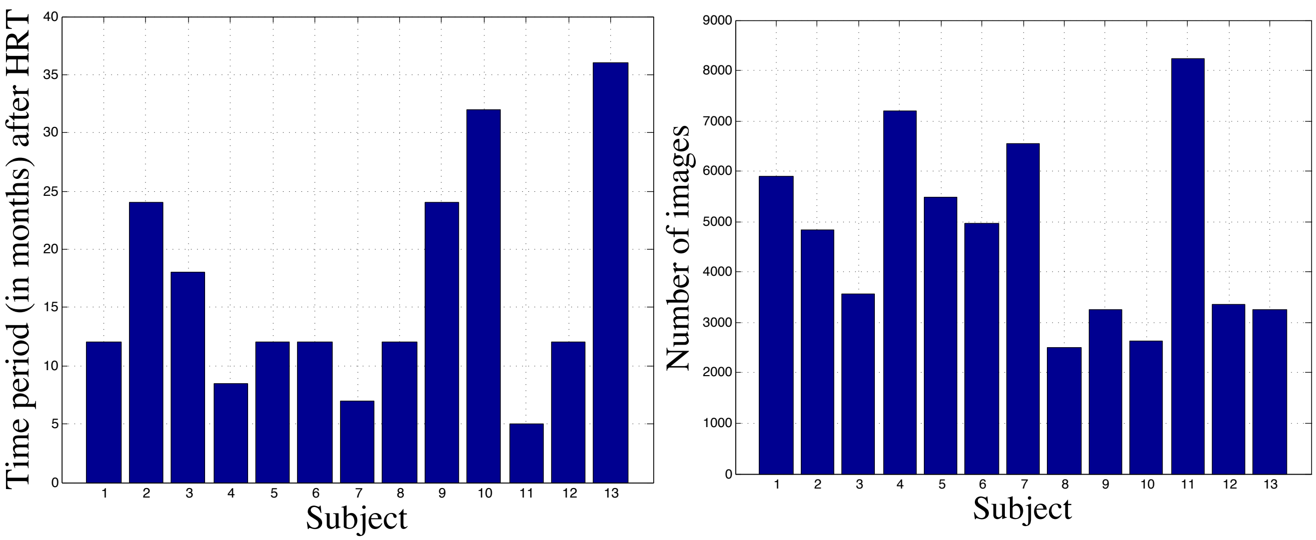
We perform our experiments on recently introduced HRT transgender dataset [3]. The dataset consists of videos belonging to 38 subjects from YouTube videos. These videos are the compilations of the images taken during various time periods of the gender transformation reported by the YouTube user. There is no knowledge on whether the subjects went through a further surgery to alter their face. We, however, consider only a subset of HRT transgender dataset for our experiments. We consider only those videos from the dataset that have user annotation mentioning the time-period elapsed since the therapy. The users label the images in the video with different kinds of annotations. For example, as shown in the Figure 4, videos contain images for every day for a year, or every month or random period after the the therapy. We consider the annotations in months. For example, for all the images with labels 1- 30 days after the therapy, we gave a label as 1 month. Our subset of dataset consists of 13 subjects that have user annotations. The available annotation for each of these 13 subjects is shown in the Figure 5 (a). The videos contain images that span up to a maximum of 36 months after the therapy. In Figure 5 (b), we show the number of images available for each subject. Our final subset of dataset consists of a total 29475 images belonging to 13 subjects.

### B.Evaluation Protocol

In this section, we describe the experimental protocols used to evaluate the proposed framework for the transgender face recognition. We divide the whole database of 13 subjects that consists of 29475 images into two independent partitions namely: training and testing set. The training set consists of 900 images that represent 13 subjects at different time instance starting from 1 month to 36 months. The testing set is comprised of 28575 images corresponding to 13 subjects. The training and testing set are independent, and there is no overlapping between them. We repeat the partition of training and testing samples for $k = 10$ times and report the average result along with the standard deviation. In this work, we present the result in terms of closed identification accuracy (rank 1) or recognition rate. The closed identification rate (rank 1) is obtained by comparing 1: N subjects in the dataset, therefore, a higher the value of identification rate corresponds to the better accuracy of the algorithm.


(a) (b) (c) (d)

Fig. 4: Different kinds of annotations available in the dataset. Different subjects after (a) 12 months (b) 1 month (30 days) and (c) 1 month, from the date of therapy. (d) Some videos contain images uploaded every day for a period of one year as indicated in the beginning of the video. We have blocked the face of the user to protect their identity.




(a) (b)

Fig. 5: Statistics of HRT database (a) The annotations indicating the time elapsed since the therapy in our subset of HRT database with 13 subjects. Annotations are available from a minimum of 1 day to a maximum of 36 months after the therapy (b) Number of images available for each of the 13 subjects selected from the HRT database

## C. Results and discussion

In this section, we present the results of the proposed framework based on the Hidden Factor Analysis (HFA). To this extent, we have evaluated three different feature extraction schemes namely: LBP, HoG, and D-SIFT. Furthermore in order to have a comprehensive comparison of the proposed scheme, we compare the performance of the proposed scheme with three different classifiers namely Nearest Neighbor (1-NN), Sparse Representation Classifier (SRC) [12] and Collaborative Representation Classifier (CRC) [13].

Table I shows the quantitative results of the proposed scheme on the HRT transgender database. It is interesting to observe that the proposed framework has emerged as the best method with all three different types of feature extraction schemes. The proposed framework with LBP

TABLE I: Face recognition accuracies (mean±std-dev%) of various methods using different features.

| Method | LBP | HoG | D-SIFT | Feature level Fusion |
|---|---|---|---|---|
| 1-NN | 75.12±0.7 | 75.57±0.6 | 71.02±1.2 | 76.35±0.4 |
| SRC [12] | 70.82±0.3 | 69.86±0.5 | 64.83±1.1 | 71.86±0.4 |
| CRC [13] | 69.18±0.4 | 68.86±0.6 | 62.21±0.9 | 70.67±0.5 |
| Proposed approach | **80.01±0.4** | **81.25±0.5** | **78.66±0.9** | **82.36± 0.4** |

features shows the improvement over an average of 5% when compared with all three state-of-the-art schemes. While the proposed framework with HoG features shows an improvement of over 6% when compared with the second best method based in 1-NN. A similar improvement is also noted with the D-SIFT features that indicate an average improvement of over 7% when compared to the second best method based on 1-NN. These quantitative results show not only the superior performance of the proposed scheme but also its robustness on the transgender face recognition. Finally, we also perform the feature level fusion of all three feature extraction scheme used in this work to further improve the accuracy of the proposed framework. Thus, as noticed from the Table I, the proposed framework shows the outstanding result with a recognition rate of 82.36%. These quantitative results indicate the applicability of the proposed scheme for transgender face recognition.

Figure 6 shows the variation of the recognition accuracy of the proposed framework with various PCA dimensions, when HoG features are used. It can be observed here that the accuracy of the proposed framework is constant after a dimension of 160 features. It is also observed that the proposed framework requires a higher dimension (or number) of features when compared to that of other comparison schemes at the cost of improved recognition accuracy. However, the features dimension of 160 features is quite reasonable to achieve real time performance.

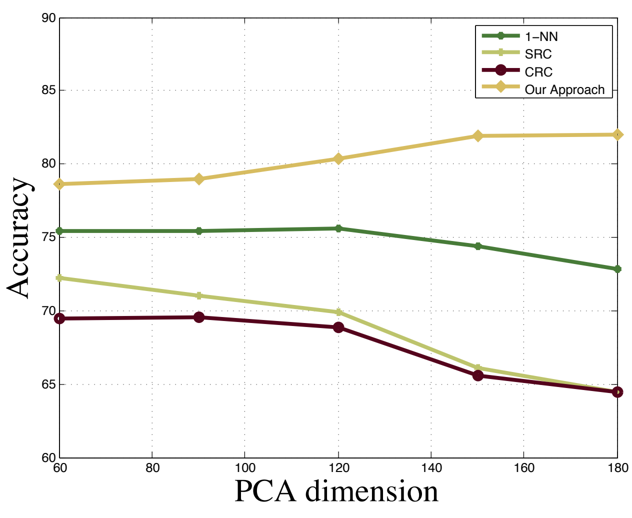



Fig. 6: Face recognition accuracies of various algorithms in (%) for different PCA dimension using HOG features.

Figure 7 shows the variation of the recognition accuracy with number of training samples at each time period. It is quite interesting to observe that, the performance of the proposed framework is quite stable after 3 training images. This further shows the applicability of the proposed framework for transgender face recognition application.

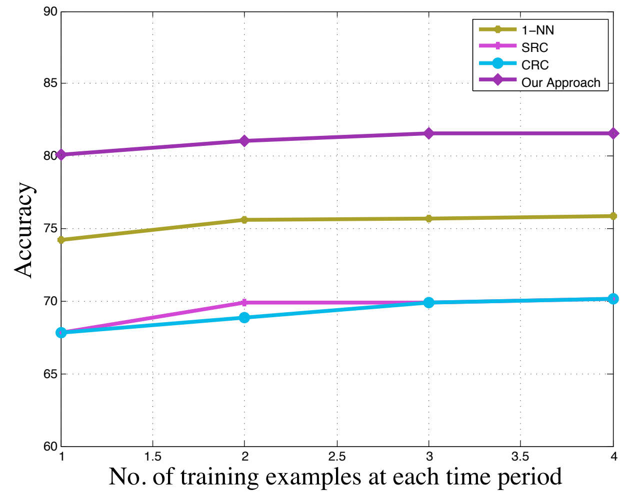



Fig. 7: Performance of various face recognition algorithms with varying number of training samples

Thus, based on the extensive experiments reported above the proposed framework based on the Hidden Factor Analysis (HFA) has emerged as the best method for transgender face recognition. When compared to the existing methods that are based on the eye region [4] and the component based approach that include periocular, nose and mouth region from the face [5], the proposed framework has the following advantages:

- Overcomes the need of the face components (like periocular, nose and mouth ) that not only difficult to extract with unconstrained face images but also requires additional computation as they need eye, nose and mouth detectors to localize and process the extracted components from the face image.
- Capable of handling low quality face images (see Figure 2) that are captured in unconstrained conditions.

We could not present the comparison result of the proposed approach with existing schemes [4], [5] because of the non-availability of the periocular regions in case of [4] and face component regions in case of [5] that are manually extracted [1]. Furthermore, the proposed framework is based on a learning scheme that will learn both appearance and the therapy component in contrary to the existing methods [4], [5] that ignore such therapy changes.

[1]Authors have tried to use the Haar cascade detector to detect the periocular and other components of the face including nose and mouth. However, due to the low quality images in HRT database, the detection accuracy is not enough to re-implement and compare the existing schemes.

## V. Conclusion

Transgender face recognition is the emerging challenge in the field of face biometrics. In this paper, we have proposed a novel framework by exploring both appearance and transformation factors using Hidden Factor Analysis (HFA). We propose a solution based on Hidden Factor Analysis (HFA) to jointly model a face under therapy as a linear combination of appearance and transformation factors. We conduct extensive experiments publicly available HRT transgender face database and demonstrate the superiority of our HFA model over state-of-the-art schemes.

## VI. acknowledgment

This work is carried out under the funding of the Research Council of Norway (Grant No. IKTPLUSS 248030/O70).

## References

[1] C. Chen, A. Dantcheva, and A. Ross. Automatic facial makeup detection with application in face recognition. In *Biometrics (ICB), 2013 International Conference on*, pages 1–8, June 2013.

[2] D. Gong, Z. Li, D. Lin, J. Liu, and X. Tang. Hidden factor analysis for age invariant face recognition. In *IEEE International Conference on Computer Vision, ICCV*, 2013.

[3] G. Mahalingam, Karl Ricanek Jr., and A. M. Albert. Investigating the periocular-based face recognition across gender transformation. *IEEE trans. on Information Forensics and Security*, 2014.

[4] G. Mahalingam and K. Ricanek. Is the eye region more reliable than the face? a preliminary study of face-based recognition on a transgender dataset. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–7, Sept 2013.

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

[6] R. Raghavendra and C. Busch. Novel presentation attack detection algorithm for face recognition system: Application to 3d face mask attack. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 323–327, 2014.

[7] R. Raghavendra and C. Busch. Robust 2d/3d face mask presentation attack detection scheme by exploring multiple features and comparison score level fusion. In *17th International Conference on Information Fusion (FUSION-2014)*, pages 1–7, 2014.

[8] R. Raghavendra, B. Yang, and C. Busch. Accurate face detection in video based on likelihood assessment. *Norsk informasjonssikkerhetskonferanse (NISK)*, 2012.

[9] R. Singh, M. Vatsa, H. Bhatt, S. Bharadwaj, A. Noore, and S. Nooreyezdan. Plastic surgery: A new dimension to face recognition. *Information Forensics and Security, IEEE Transactions on*, 5(3):441–448, Sept 2010.

[10] A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. 2010.

[11] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154, 2004.

[12] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust Face recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2009.

[13] L. Zhang, M. Yang, and X. Feng. Sparse representation or Collaborative representation: Which helps face recognition? *IEEE International Conference on Computer Vision, (ICCV)*, 2011.