2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Face Recognition for Newborns, Toddlers, and Pre-School Children: A Deep Learning Approach

Sahar Siddiqui, Mayank Vatsa, and Richa Singh
IIIT-Delhi, India
{sahar14091, mayank, rsingh}@iiitd.ac.in

***Abstract*—Biometric recognition of newborns, toddlers, and pre-school children aims is an important research challenge with applications in identifying newborn swapping, missing kids, and disbursing benefits. In this research, we propose a representation learning algorithm to extract unique and invariant features from face images of newborns and toddlers, to design an efficient face recognition algorithm. Specifically, we propose a deep learning model which applies class-based penalties while learning the filters of a convolutional neural network. The proposed CNN architecture achieves a rank-1 identification accuracy of 62.7% for single gallery newborn face recognition and 85.1% for single gallery toddler face recognition, forming state-of-the-results for both the databases. Comparison with several existing algorithms also showcases the effectiveness of the proposed algorithm on both the databases.**

I. Introduction

Biometric modalities are nowadays used in several identity management applications such as national identification projects, law enforcement, access control, and surveillance. However, there is very limited literary evidence to prove the reliability of these modalities for recognition of newborns and toddlers, especially in the age group of 0-4 years. In a recent incident in India [1], 14 children were rescued from places where they were found begging and picking rags as part of forced labor. However, due to lack of biometric records for these children, there was no way of identifying their birth parents or verifying the claims of people who demanded the custody of those children. The District Child Protection Officer who was in-charge urged the need for biometric identification of children who might get abducted from far off places and forced into begging and child labor. At present, the Unique Identification Authority of India (UIDAI) [2] does not record any biometric information of children below the age of 5 years. In a country like India where about 9.7% of the population belongs to the age group of 0-4 years [3], it becomes essential to identify every child with their biometric identities uniquely.

One of the major challenges lies in identifying the biometric modality best suited for this age group (0 to 4 years). In the past few years, studies have been conducted to determine the reliability of fingerprints [4], footprints [5], palm prints [6], face [7], [8], [9], iris [10], and a fusion of some of these modalities [10] for biometric recognition of newborns and toddlers separately. Studies have shown that most of these modalities including fingerprint and iris [11], in its current form, may not be useful for effective identification of newborns and toddlers. The primary reasons for this are

Figure 1: Illustrating the challenges involved with face recognition for newborns, toddlers, and pre-school children. The images are taken from the Newborn Face Database [8] and Children Multimodal Biometric Database [10].

the inability to capture data (for instance, unable to open eyes right after birth for iris recognition and skin elasticity in fingerprints), temporal variations, and reliability of features.

This research focuses on face recognition for newborns and toddlers. As shown in Figure 1, recognizing faces of newborns and toddlers is a challenging problem due to the *unintentionally uncooperative behavior* of children while capturing face images. Moreover, limited availability of such datasets has also hindered research in this domain.

A. Related Work

Biometric identification of children has been studied sporadically but it started as early as 1899 when Galton *et al.* [12] collected inked fingerprint impression of a single child and manually evaluated the feasibility of fingerprint recognition. Recently, Jain *et al.* [4] and Basak *et al.* [10] have shown the plausibility of using fingerprint and iris recognition for newborns and toddlers, respectively. With respect to face modality, in a preliminary study, Bharadwaj *et al.* [7] articulated the challenges of newborn face recognition. As a preliminary approach, they presented an algorithm comprising multi-resolution texture representations from three scales for effective newborn face matching. In another study, Bharadwaj *et al.* [8] proposed a two-stage domain-specific learning for newborn face recognition. Basak *et al.* [10] produced baseline results for toddler face recognition by evaluating existing tools and algorithms. In addition to providing the preliminary approaches, Bharadwaj *et al.* [8] and Basak *et al.* [10] also

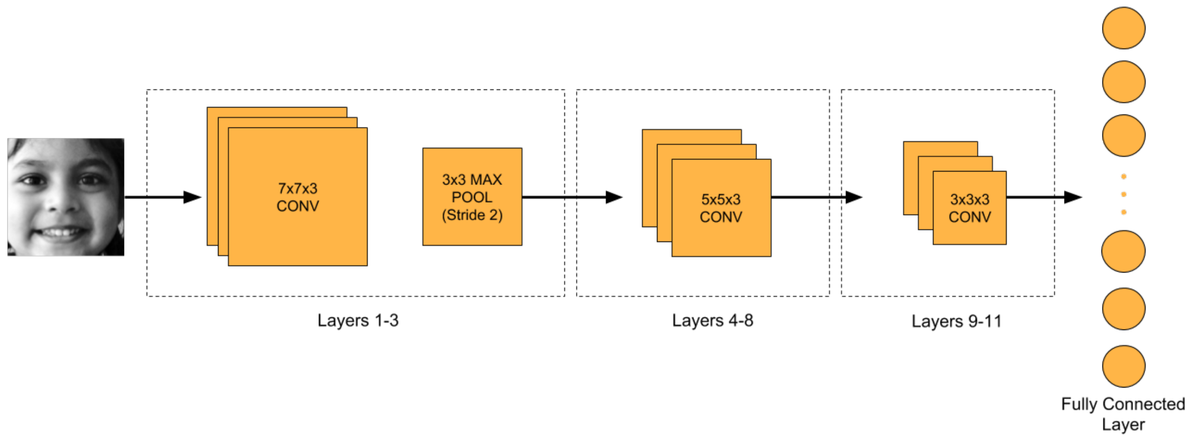


Figure 2: Steps involved in the proposed CNN feature extraction with class-based penalty for filter learning. For matching, the classifier is added separately depending on the identification or verification experiments.

introduced the Newborn Face Database and Children Multi-modal Biometric Database (CMBD), respectively. Availability of these databases has further motivated us to continue the research in this domain and establishing the reliability of face as an effective modality for unique identification of newborns and toddlers. Recently, Jain *et al.* [9] also introduced the Newborns, Infants, and Toddlers Longitudinal (NITL) face image database and demonstrated the results of commercial-off-the-shelf (COTS) face matcher on this database. However, this database is not publicly available.

B. Research Contributions

The major contribution of this research lies in designing a face recognition algorithm for both newborns and toddlers, which achieves state-of-the-art results with a single gallery image. Specifically,

- a novel approach is proposed for inducing class-based penalty to learn the filters in Convolutional Neural Network (CNN),
- the proposed model is utilized to design a face recognition algorithm for efficiently matching images of newborn, toddlers, and pre-school children, and
- the results of the proposed algorithm are demonstrated on the Extended Newborn Face Database and Children Multimodal Biometric Database [10].

Next section describes the proposed face recognition algorithm. Section III presents the databases and algorithms used for evaluating and comparing the performance of the proposed algorithm. The experimental results are discussed in Section IV followed by conclusion in Section V.

II. Proposed Face Recognition Algorithm

CNNs have gained a lot of popularity in the computer vision community and led to significant improvements in state-of-the-art results of many applications including face recognition [13]. In this research, we present a modification in the convolutional filter learning and impose a class-based penalty on the weights of the filters. This section explains the proposed CNN model and its implementation details for face recognition for newborn, toddlers and pre-school children.

A. Class based Penalty in Filter Learning

For a C class problem, let X be the input to a convolutional layer with N filters. Let W_i and b_i be the weight matrix and bias of the i^{th} filter, respectively. The process of learning feature maps in CNN is shown in Equation 1.

$$F_i = \sigma\left(W_i X + b_i\right) \quad \forall i = 1 \dots N \tag{1}$$

where, σ is the activation function such as ReLU ($\sigma(z) = \max(0, z)$). For pair-wise input, we extend Equation 1 by incorporating a class-based penalty term, i.e.,

$$F_i = \sigma\left(W_i X + b_i\right) \oplus \left(\lambda \left\|\mathcal{C}_i W_i\right\|_2^2 + \beta\left(T_r\left(H_i^T H_i L\right)\right)\right) \tag{2}$$

where, $\oplus$ is a special operator which symbolizes that the class-based penalty is added to the weight matrix of the i^{th} filter. Here, λ and β are the regularization constants, and $\mathcal{C}_i$ is the set of weights associated with each class. Further, T_r denotes the trace of a matrix and $H_i = \sigma\left(W_i X + b_i\right)$. L is the Laplacian matrix constructed as $L = D - M$, where $D = diag\left(d_1, d_2 \dots d_s\right)$, and $d_k = \sum_{j=1}^{c} M_{kj}$. Finally, M is defined as,

$$M_{kj} = \begin{cases} +1 \text{ if } x_k \text{ and } x_j \text{ are of the same class} \\ -1 \text{ if } x_k \text{ and } x_j \text{ are of different classes} \end{cases} \tag{3}$$

Using Equation 2, we learn discriminative (supervised) feature maps with class-based penalty imposed on the weights of the filters. Since both the additional terms in Equation 2 are convex and smooth, it is differentiable, and Adam optimizer with cross-entropy as the loss function can be applied to learn the filters. The class-based penalty helps in learning class-specific discriminative features and avoid over-fitting. Further, in order to learn higher level features with increasing number of layers, the network is trained with skip residual connections [14].

Figure 3: Samples from the Extended Newborn Database [8].

Figure 4: Samples from the Children Multimodal Biometric Database [10].

B. Incorporating the Proposed CNN Architecture for Newborn, Toddler, and Pre-school Children Recognition

For recognizing newborn, toddler, and pre-school children, 11-layer network is formed using the proposed architecture. Figure 2 illustrates the feature extraction process using the proposed architecture. The input to the network is a 200×200 sized face image. In the first three layers of CNN, three filters of size 7×7 are used, followed by max pool with size 3×3 and stride size 2. In the next five layers, three filters of size 5×5 are used, and in the last three layers, three filters of size 3×3 are used. Finally, a flattened fully connected layer followed by softmax is used for classification. ReLU is used as the activation function. Batch normalization, as well as standard data augmentation techniques, are used during training of the network.

Since the size of the newborn and toddler training set is small, we have used CMU-MultiPIE database [15] to pre-train the network. Furthermore, during fine tuning on the newborn and children databases, data augmentation techniques have been applied to increase the size of training sets of respective databases. The deep learning model is trained in the verification mode (1:1 matching) and for identification, N-way verification is performed.

III. Databases and Experimental Protocol

This section provides the details of the databases, experimental protocol, and algorithms used for comparison.

A. Datasets

In order to evaluate the performance of all the algorithms, we have used two datasets:

- **Extended Newborns Face Database**: Bharadwaj *et al.* [8] introduce the Newborn Face Database which consists of over 800 near frontal face images of 96 newborns. The dataset is further extended to include 1185 images pertaining to 204 newborns, each having 1-17 images, collected from various hospitals. The time of capture varies from one hour to a few weeks after birth. Sample images of two babies illustrating the associated challenges are shown in Figure 3.
- **Children Multimodal Biometric Database (CMBD)** [10]: It consists of 2590 face images of 141 toddlers and pre-school children (age range of 18 months to 4 years), each having 10-20 images. Apart from face images, the database also consists of iris and fingerprint images of 100 toddlers; however, in this research, we have only used the face images. This database has been acquired over two sessions which are months apart. Figure 4 shows the sample images from the CMB database.

B. Experimental Protocol

Face images in both the datasets are detected using the Viola Jones face detector and aligned using affine transformation with the inter-eye distance set as 100 pixels. Images in which face is not automatically detected are manually processed. In order to maintain uniformity, every image used in the experiment has been resized to 200×200 pixels and pre-processed using histogram normalization.

For performance evaluation, both the databases are individually divided into two parts: training and testing, with a label-wise $60 : 40$ split. The testing set is further divided into gallery and probe sets. The first image of each individual in the testing split is selected for creating the gallery, and the remaining images are chosen as probe images. These probe images are matched with the gallery images to compute the results. The results are reported in terms of rank-1 identification accuracy and receiver operating characteristics (ROC) curve with five times random cross-validation.

C. Comparative Algorithms

In order to evaluate the effectiveness of the proposed algorithm, the comparisons are performed with the following handcrafted and learning-based feature extraction algorithms.

Subspace Approaches: Two different subspace approaches have been used: Principal Component Analysis and Linear Discriminant Analysis.

- *Principal Component Analysis* (PCA) [16] is a statistical approach used for dimensionality reduction. In face recognition, PCA is used for reducing the image variables by selecting only the features which show maximum variance across different subjects. Every image in the training set is represented as a linear combination of its eigenvectors. Based on the eigenvalues, a set of significant eigenvectors (Eigenfaces for images) are selected to represent the subspace. Features are extracted by transforming the image to this subspace by taking a dot product. Further, Cosine Similarity is used to measure the similarity between two feature vectors.

- *Linear Discriminant Analysis* (LDA) [17] works on a similar principle as PCA and is used for dimensionality reduction while preserving inter-class discriminatory information. The images are transformed by projecting them to Fisherspace, and the transformed image consists of the features extracted. Similar to PCA, Cosine Similarity is used for feature matching.

Texture Algorithm - *Local Binary Patterns*: LBP [18] descriptor is widely used in face recognition due to its property of being illumination invariant. The circular LBP descriptor, an extension to LBP, assigns a discrete value to a pixel by thresholding a window of P pixels on a circle of radius R with the center pixel value and considers the result as the binary number representation. There is no specific training algorithm in this technique. LBP descriptor directly extracts the features of the test images, and these features are compared for matching using χ^2 distance.

Deep Learning: Two off-the-shelf CNN architectures are also used for evaluation and comparison.

- *Fine-tuned VGG Face*: For the VGG-Face [13], pre-trained model based on MatConvNet implementation is used and the last six convolutional layers are fine-tuned with the Newborns' and Toddlers' face images for their respective experiments.
- *Triplet Convolutional Neural Network (Triplet CNN):* Triplet CNN is inspired by the concept of Siamese Network which learns via comparative measures rather than the labels and makes use of triplet loss as part of the deep metric learning. We have implemented our own version of the deep convolutional network as described in [19] where cosine distance is used to calculate face similarity while computing the triplet loss. In this case, the CNN used in the Triplet CNN is the fine-tuned VGG-Face. The final fully connected layer of the CNN produces a vector of 128 units. The last six convolutional layers in the VGG-Face are trained as part of the Triplet CNN using triplet loss, and the classifier is able to classify the distinguishing features extracted from the input face image for recognition.

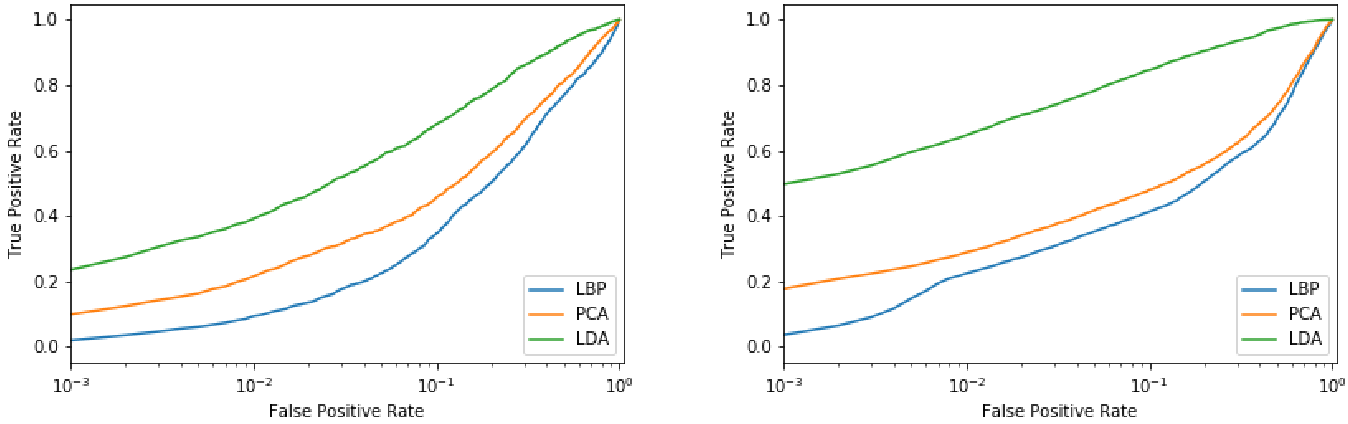
IV. Analysis of Experimental Results

The performance of the proposed deep learning algorithm is compared with existing handcrafted and learning-based techniques. The results of the proposed algorithm and existing algorithms (PCA, LDA, LBP, Fine-tuned VGG Face, and Triplet CNN) are computed on both Newborn Face Database and Children Multimodal Biometric Database. The rank-1 accuracies are summarized in Table I and the ROC curves are shown in Figure 5 and Figure 6. The key analyses are explained below:

Table I: Average Rank-1 identification accuracies and standard deviation (%) of the algorithms over five times cross validation on the Extended Newborn Face Database and Children Multimodal Biometric Database.

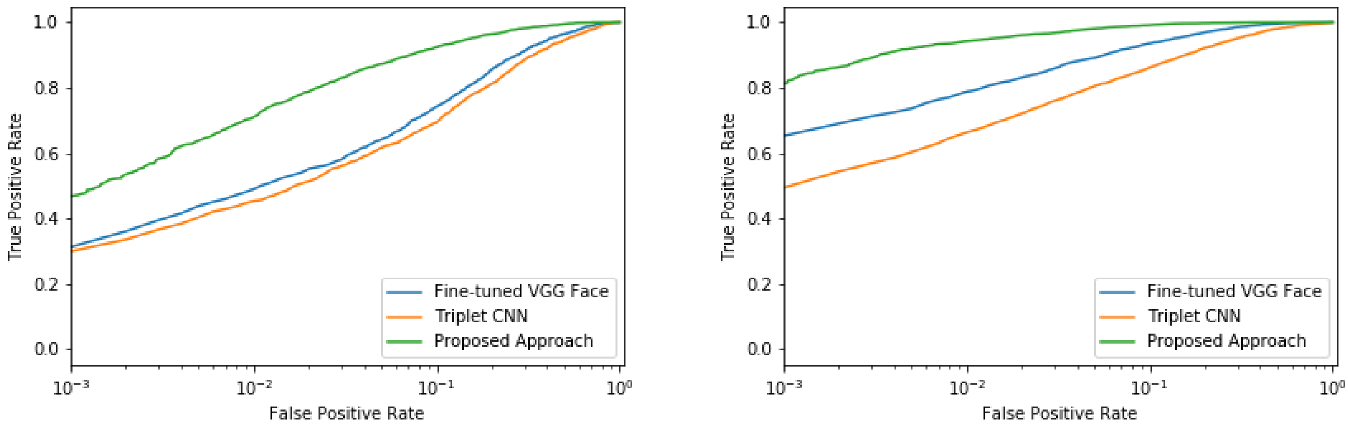
Algorithms	Newborn Database	CMB Database
PCA	24.8 $\pm$ 0.03	38.8 $\pm$ 0.03
LBP	24.5 $\pm$ 0.02	28.8 $\pm$ 0.03
LDA	43.3 $\pm$ 0.04	71.3 $\pm$ 0.02
Fine-tuned VGG-Face	54.1 $\pm$ 0.07	83.0 $\pm$ 0.04
Triplet CNN	48.8 $\pm$ 0.05	72.7 $\pm$ 0.02
Proposed	62.7 $\pm$ 0.04	85.1 $\pm$ 0.05

- Out of the existing subspace and texture based algorithms, LDA works better than the rest and yields rank-1 accuracies of 43% and 71% on the two databases. This shows that the traditional feature extraction techniques are unable to model the variations in face images of newborns and toddlers.
- All the baseline techniques have a visibly better performance on the Children Multimodal Biometric Database compared to the Newborn Face Database suggesting that the features are more discriminatory in case of toddlers. We have also observed that the expression variations in case of toddler's faces are significantly reduced compared to newborns, which further helps in improving the recognition performance.
- The results show that the deep learning algorithms perform better than the existing handcrafted techniques. Among the existing algorithms, on the two databases (with single gallery per subject), the fine-tuned VGG-Face performs the best with a rank-1 identification accuracy of 54.1% and 83%, respectively.
- The proposed approach yields significantly better results than the other deep learning models. It is important to note that VGG-Face and Triplet CNN are trained with multiple folds higher number of training images compared to the proposed approach which is pre-trained on frontal face images from the CMU MultiPIE database. The two factors for high accuracies are paired training, which increases the number of training samples, and class-based penalty which enforces class-aware filter learning.
- In face verification experiments, ROC curves on both the databases also show that the proposed algorithm consistently yields better results even at lower false accept rates. This shows the efficacy of the proposed model.
- Figure 7 shows some sample cases of correct and incorrect classifications by the proposed and existing algorithms. It can be observed that the proposed algorithm can handle the presence of multiple variations at a time, for instance, the first two rows of images contain illu-



(a) Newborn Face Database (b) Children Multimodal Biometric Database

Figure 5: ROC curves for PCA, LBP, and LDA on Newborn Face Database (left) and CMBD (right).



(a) Newborn Face Database (b) Children Multimodal Biometric Database

Figure 6: ROC curves for Fine-tuned VGG Face, Triplet CNN, and the proposed approach on the two databases.

mination, resolution, or pose variations. However, there are some cases where the proposed algorithm failed to perform well, specifically, due to extreme expression variations.

- Computationally, the proposed algorithm requires less than 0.1 seconds to match a pair of images on a Tesla K40 workstation with 128GB RAM.

V. Conclusion

This research improves state-of-the-art for matching face images of newborns, toddlers and preschool children. We first present a novel formulation to incorporate class-based penalty for filter training in the CNN architecture. The proposed approach is extended for face recognition for newborns, toddlers, and pre-school children. The publicly available newborn face database is also extended to more than double in terms of the number of subjects, and the performance is evaluated on both extended newborn database and children multimodal biometric database. The proposed deep learning approach outperforms existing handcrafted and deep learning based approaches and achieves state-of-the-art verification and identification accuracies for both the databases.

Moving forward, we plan to work on video based face recognition for recognizing face images pertaining to newborns, toddlers and preschool children. Algorithms such as frame selection in videos [20], joint-feature learning [21], and fusion [22] can help improve the performance of face recognition in realistic conditions.

VI. Acknowledgement

The authors acknowledge C. Parmar for her help in extending the Newborn Face Database. M. Vatsa and R. Singh are partially supported through Infosys Center for Artificial Intelligence, IIIT-Delhi, India.

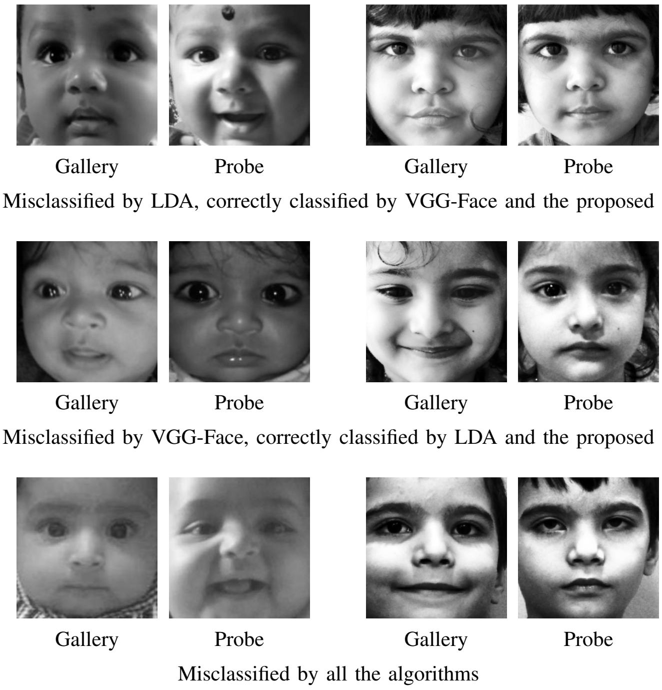


Figure 7: Analyzing the results of the proposed algorithm and comparison with existing algorithms.

References

[1] *http://www.tribuneindia.com/news/amritsar/now-biometric-identification-for-child-beggars-parents/530244.html*, Jan 18, 2018.

[2] UIDAI. *https://uidai.gov.in/*.

[3] Census of India. *http://www.censusindia.gov.in/vital_statistics/srs_report/9chap2-2011.pdf*, 2011.

[4] A. K. Jain, S. S. Arora, K. Cao, L. Best-Rowden, and A. Bhatnagar. Fingerprint recognition of young children. *IEEE Transactions on Information Forensics and Security*, 12(7):1501–1514, 2017.

[5] D. Weingaertner, O. Bellon, L. Silva, and M. N. L. Cat. Newborn's biometric identification: Can it be done? In *3rd International Conference on Computer Vision Theory and Applications*, volume 1, pages 200–205, 2008.

[6] W. Jia, H. Cai, J. Gui, R. Hu, Y. Lei, and X. Wang. Newborn footprint recognition using orientation feature. *Neural Computing and Applications*, 21(8):1855–1863, 2012.

[7] S. Bharadwaj, H. S. Bhatt, R. Singh, M. Vatsa, and S. K. Singh. Face recognition for newborns: A preliminary study. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6, 2010.

[8] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh. Domain specific learning for newborn face recognition. *IEEE Transactions on Information Forensics and Security*, 11(7):1630–1641, 2016.

[9] L. Best-Rowden, Y. Hoole, and A. Jain. Automatic face recognition of newborns, infants, and toddlers: A longitudinal evaluation. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–8, 2016.

[10] P. Basak, S. De, M. Agarwal, A. Malhotra, M. Vatsa, and R. Singh. Multimodal biometric recognition for toddlers and pre-school children. *IEEE International Joint Conference on Biometrics*, 2017.

[11] Aadhar card: Registration for newborns starts in two state hospitals. https://tinyurl.com/y7ep5596.

[12] F. Galton. Finger prints of young children. *Report of the British Association for the Advancement of Science*, 69:868–869, 1899.

[13] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[15] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. In *IEEE International Conference on Automatic Face Gesture Recognition*, pages 1–8, 2008.

[16] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[17] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[18] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.

[19] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[20] G. Goswami, R. Bhardwaj, R. Singh, and M. Vatsa. Mdlface: Memorability augmented deep learning for video face recognition. In *IEEE International Joint Conference on Biometrics*, pages 1–7, 2014.

[21] G. Goswami, M. Vatsa, and R. Singh. Face verification via learned representation on feature-rich video frames. *IEEE Transactions on Information Forensics and Security*, 12(7):1686–1698, 2017.

[22] H. S. Bhatt, R. Singh, and M. Vatsa. On recognizing faces in videos using clustering-based re-ranking and fusion. *IEEE Transactions on Information Forensics and Security*, 9(7):1056–1068, 2014.