

Learning A Shared Transform Model for Skull to Digital Face Image Matching

Maneet Singh¹, Shruti Nagpal¹, Richa Singh¹, Mayank Vatsa¹, and Afzel Noore²

¹IIT-Delhi, India, ²Texas A&M University, Kingsville, USA

{maneets, shrutin, rsingh, mayank}@iitd.ac.in, afzel.noore@tamuk.edu

Abstract

Human skull identification is an arduous task, traditionally requiring the expertise of forensic artists and anthropologists. This paper is an effort to automate the process of matching skull images to digital face images, thereby establishing an identity of the skeletal remains. In order to achieve this, a novel Shared Transform Model is proposed for learning discriminative representations. The model learns robust features while reducing the intra-class variations between skulls and digital face images. Such a model can assist law enforcement agencies by speeding up the process of skull identification, and reducing the manual load. Experimental evaluation performed on two pre-defined protocols of the publicly available IdentifyMe dataset demonstrates the efficacy of the proposed model.

1. Introduction

Human skull identification is a challenging problem of vital importance in forensics and law enforcement, which involves matching a skull image with gallery face images (Figure 1). Unidentified skulls often result in unrest with the law enforcement officials and family of the victim. For instance, in 2014, skeletal remains (including a skull) were found in North Creek, just above Virginia [15]. Forensic experts were able to predict an age-range, gender, and ethnicity of the skull, however, the identity remains unknown. Efforts were made to investigate missing person reports around that area, and other unsolved cases. Retired police chief, Kurt Wright, who worked on the case, mentioned “This has always just kind of nagged at me for some reason, and I wanted some closure to it” [15]. In 2017, the case was again publicized in news to look for fresh leads; however, the case remains unsolved.

As shown in Figure 2, skull identification is primarily performed by means of a manual reconstruction approach. That is, given a skull image, a facial reconstruction is created which is then circulated via digital or print media, or matched with a given gallery of digital face images. The techniques for facial reconstruction can broadly be divided

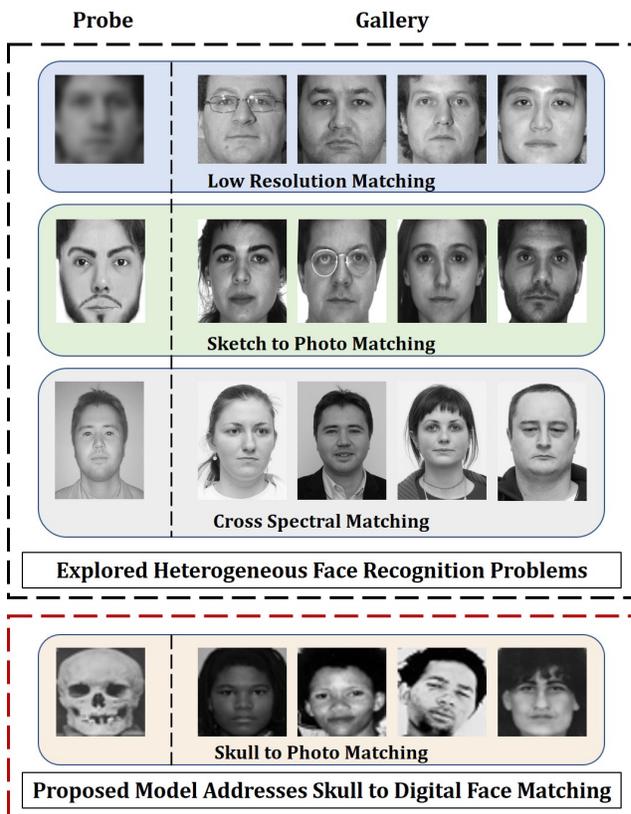


Figure 1: In literature, heterogeneous matching problems such as low resolution, sketch face recognition, and visible to near infrared (cross spectral) face recognition have received substantial attention. This research addresses the less explored heterogeneous problem of skull to digital face image matching.

into three categories: (i) 2-D reconstruction, (ii) 3-D reconstruction, and (iii) superimposition. 2-D reconstruction techniques construct a 2-D face image from a given skull image, whereas 3-D reconstruction constructs a 3-D face, using either Computer Graphics or via clay or sculptures. Superimposition refers to the task of superimposing a digital face image over a skull, and analyzing its alignment ac-

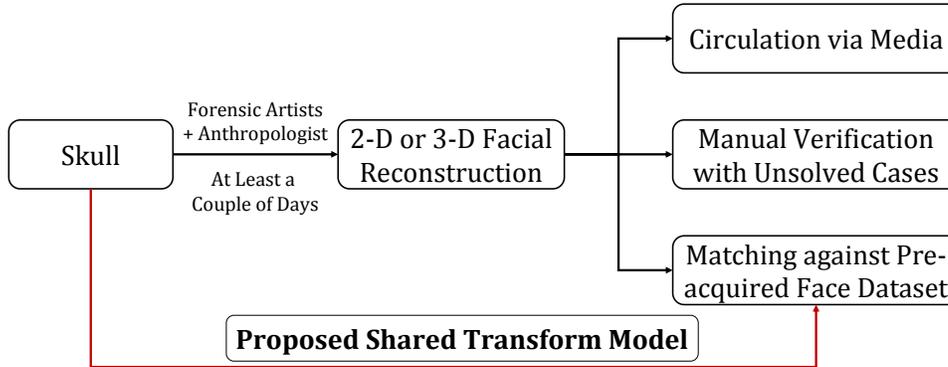


Figure 2: Traditional techniques utilize facial reconstruction, a time consuming process, often requiring the input of interdisciplinary experts. The proposed Shared Transform Model eliminates the need for reconstruction by directly matching a skull image with a dataset of face images.

curacy. This technique is generally applied when the forensic experts have a small subset of possible identities to be verified against a given skull. Traditionally, the successful application of the above techniques require at least a couple of days to a couple of weeks, along with collaboration between forensic artists and anthropologists. In this research, we propose to automate the process of skull identification. If a completely intact skull is found, it may enable the law enforcement agencies to determine its identity without performing the facial reconstruction. With the availability of a skull recognition algorithm, the search for the identity of a skull can extend beyond the immediate location in a fast, effective, and automated manner.

Skull identification can be viewed as a heterogeneous matching problem, where one domain corresponds to skull images, while the second contains digital face images. Figure 1 shows different heterogeneous face matching problems explored in literature along with a sample skull to digital face matching scenario. It can be observed that the information content of digital face images varies significantly from that of the skull image. Heterogeneous face matching has received significant attention in literature, in terms of low resolution face matching, or sketch to face matching, or visible to near infrared (NIR) face matching [10, 12]. However, the task of automated skull recognition has received limited attention [1, 4]. Most of the existing algorithms focus on the task of facial reconstruction [6] by means of superimposition, that is, scenarios having the availability of a small subset of possible identities for a given skull. Besides reconstruction, researchers have also focused on automating the task of matching 3-D CT head scans with a gallery of digital images [5, 16]. While this is a viable option to perform matching, it incurs the overhead charge of conducting a CT head scan, and the inherent assumption that the skull found can be used for scanning. A model

for automated skull recognition without the need of reconstructions or scans has been proposed by Nagpal *et al.* [8]. The authors also proposed a publicly available IdentifyMe dataset consisting of 464 skull images, along with Semi-supervised transform learning (SS-TL) and Unsupervised transform learning (US-TL) models for skull to digital face image matching.

Existing algorithms for skull recognition rely either on the availability of CT head scans or require learning a large number of parameters. Given the availability of limited training samples, this might result in over-fitting and thus lower the identification performance on unseen probes. In order to address these limitations, this research proposes a novel Shared Transform Model (Figure 3), which learns a single “shared transform” for both skulls and digital face images, while reducing the intra-class variations. Experimental results on the publicly available skull dataset, IdentifyMe [8] demonstrate the efficacy of the proposed model for the task of skull to face matching.

2. Proposed Shared Transform Model

Figure 4 presents sample skull and digital face image pairs. It can be observed that the images corresponding to a given subject showcase large variations in terms of the information content. Coupled with limited training samples, high intra-class variations render the problem of skull to digital face image matching challenging. Existing research in skull to photo matching has utilized transform learning [8]. The superior performance of transform learning models for other heterogeneous tasks with limited training samples [9] further make it an appropriate choice for addressing the given problem. However, the models proposed by Nagpal *et al.* [8] for the given task, SS-TL and US-TL, require learning a lot of parameters and include separate transforms for skulls and faces. In order to reduce the number

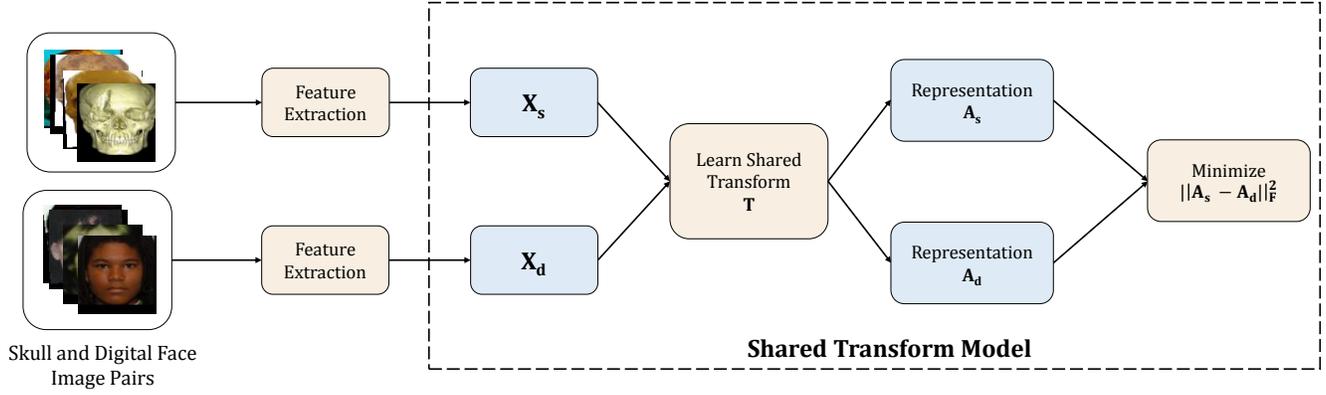


Figure 3: Diagrammatic representation of the proposed Shared Transform Model. Given mated pairs of skull and digital face image (\mathbf{X}_s and \mathbf{X}_d), the proposed algorithm learns a shared transform matrix \mathbf{T} , along with the corresponding representations \mathbf{A}_s and \mathbf{A}_d . The model is optimized in order to reduce the intra-class variations of the given training data.

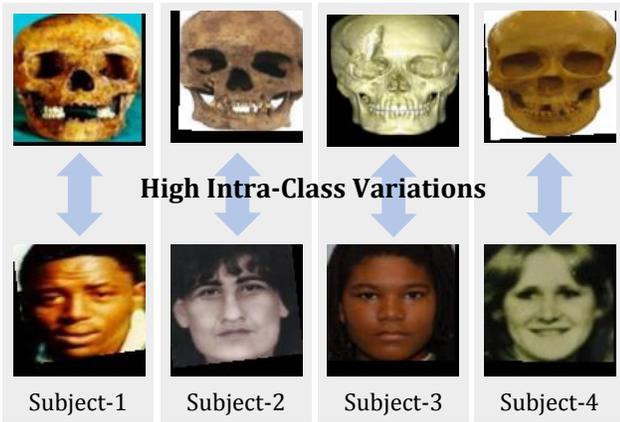


Figure 4: Sample images of four subjects of the IdentifyMe dataset. The top row corresponds to the skull images, while the bottom rows contains the digital face images. Large variations can be observed across the two images of a particular subject.

of parameters and improve the overall recognition performance, in this research, a Shared Transform Model is proposed for learning representations such that the intra-class variations are minimized. Learning a shared transform reduces the number of parameters to be learned, and minimizing the intra-class variations incorporates discriminability in the proposed model.

Transform Learning algorithms [13, 14] learn a transform matrix \mathbf{T} for an input \mathbf{X} , in order to generate a sparse representation \mathbf{A} . Mathematically, this is represented as:

$$\min_{\mathbf{T}, \mathbf{A}} \|\mathbf{TX} - \mathbf{A}\|_F^2 + \lambda_1 \|\mathbf{T}\|_F^2 + \lambda_2 \log \det(\mathbf{T}) \quad (1)$$

$$s.t. \|\mathbf{A}\|_0 \leq \tau$$

where, the second and the third terms correspond to regu-

larizers for ensuring a non-degenerate and scale balanced transform matrix. λ_1 and λ_2 control the weight given to the regularizers while optimizing the entire loss function. The sparsity constraint enforces the model to learn representative latent variables, thereby extracting meaningful information from the given data.

The proposed Shared Transform Model learns representations such that the features of skull and digital face images are projected onto a common space, while reducing the intra-class variations. A shared transform is learned by projecting the given data onto a common space. That is, using the training data: skulls (\mathbf{X}_s) and digital face images (\mathbf{X}_d), a shared transform (\mathbf{T}) is learned as follows:

$$\min_{\mathbf{T}, \mathbf{A}_d, \mathbf{A}_s} \|\mathbf{TX}_s - \mathbf{A}_s\|_F^2 + \|\mathbf{TX}_d - \mathbf{A}_d\|_F^2$$

$$+ \lambda_1 \|\mathbf{T}\|_F^2 + \lambda_2 \log \det(\mathbf{T}) \quad (2)$$

$$s.t. (\|\mathbf{A}_s\|_0 \leq \tau, \|\mathbf{A}_d\|_0 \leq \tau)$$

Equation 2 learns a shared transform matrix \mathbf{T} and sparse representations (\mathbf{A}_s , \mathbf{A}_d) for skull and digital face images, respectively. This projects the input data onto a common feature space; however, it does not ensure that the intra-class variations are minimized. In order to obtain features that reduce the within-class variations, the proposed Shared Transform Model incorporates a distance minimizing term between the representations as follows:

$$\min_{\mathbf{T}, \mathbf{A}_d, \mathbf{A}_s} \|\mathbf{TX}_s - \mathbf{A}_s\|_F^2 + \|\mathbf{TX}_d - \mathbf{A}_d\|_F^2$$

$$+ \lambda_1 \|\mathbf{T}\|_F^2 + \lambda_2 \log \det(\mathbf{T}) + \lambda_3 \|\mathbf{A}_d - \mathbf{A}_s\|_F^2 \quad (3)$$

$$s.t. (\|\mathbf{A}_s\|_0 \leq \tau, \|\mathbf{A}_d\|_0 \leq \tau)$$

Similar to Equation 2, the above objective function learns a shared transform matrix \mathbf{T} , along with the representations of skull and digital face images (\mathbf{A}_s , \mathbf{A}_d). The

additional distance minimizing term, $(\|\mathbf{A}_d - \mathbf{A}_s\|_F^2)$ promotes learning representations such that the distance between them is minimized. Learning a shared transform matrix projects the heterogeneous input data onto a common feature space, while the distance minimizing term facilitates learning of discriminative features. Given mated pairs of skull and digital face images, the proposed Shared Transform Model can be learned via a standard alternating minimization approach which alternates over the variables to be learned for a given model. For the proposed model, this is achieved by iterating over a three step approach given below:

1. Optimize for \mathbf{T} :

$$\begin{aligned} & \min_{\mathbf{T}} \|\mathbf{T}\mathbf{X}_s - \mathbf{A}_s\|_F^2 + \|\mathbf{T}\mathbf{X}_d - \mathbf{A}_d\|_F^2 \\ & \quad + \lambda_1 \|\mathbf{T}\|_F^2 + \lambda_2 \log \det(\mathbf{T}) \\ \equiv & \min_{\mathbf{T}} \left\| \mathbf{T} \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_d \end{pmatrix} - \begin{pmatrix} \mathbf{A}_s \\ \mathbf{A}_d \end{pmatrix} \right\|_F^2 + \lambda_1 \|\mathbf{T}\|_F^2 \\ & \quad + \lambda_2 \log \det(\mathbf{T}) \end{aligned} \quad (4)$$

The above equation is of the form given in Equation 1, and is thus optimized with a closed form solution explained earlier by Ravishankar *et al.* [14].

2. Optimize for \mathbf{A}_s :

$$\begin{aligned} & \min_{\mathbf{A}_s} \|\mathbf{T}\mathbf{X}_s - \mathbf{A}_s\|_F^2 + \lambda_3 \|\mathbf{A}_d - \mathbf{A}_s\|_F^2 \\ & \quad s.t. \|\mathbf{A}_s\|_0 \leq \tau \\ \equiv & \min_{\mathbf{A}_s} \left\| \begin{pmatrix} \mathbf{T}\mathbf{X}_s \\ \sqrt{\lambda_3} \mathbf{A}_d \end{pmatrix} - \begin{pmatrix} \mathbf{I} \\ \mathbf{I} \end{pmatrix} \mathbf{A}_s \right\|_F^2 s.t. \|\mathbf{A}_s\|_0 \leq \tau \end{aligned} \quad (5)$$

3. Optimize for \mathbf{A}_d :

$$\begin{aligned} & \min_{\mathbf{A}_d} \|\mathbf{T}\mathbf{X}_d - \mathbf{A}_d\|_F^2 + \lambda_3 \|\mathbf{A}_d - \mathbf{A}_s\|_F^2 \\ & \quad s.t. \|\mathbf{A}_d\|_0 \leq \tau \\ \equiv & \min_{\mathbf{A}_d} \left\| \begin{pmatrix} \mathbf{T}\mathbf{X}_d \\ \sqrt{\lambda_3} \mathbf{A}_s \end{pmatrix} - \begin{pmatrix} \mathbf{I} \\ \mathbf{I} \end{pmatrix} \mathbf{A}_d \right\|_F^2 s.t. \|\mathbf{A}_d\|_0 \leq \tau \end{aligned} \quad (6)$$

The above two equations are of the standard form given in Equation 1, and thus have a closed form solution provided by Ravishankar *et al.* [14].

Given mated pairs of skulls and digital face images, the above three steps are optimized iteratively to obtain the final transform matrix and corresponding representations.

2.1. Identification using Shared Transform Model

The training data is used to learn the shared transform matrix (\mathbf{T}), which is utilized at the test time to perform



Figure 5: Sample skull images of the IdentifyMe dataset.

identification. In a real world scenario, the task is to match a given probe skull image (x_{Skull}) against a gallery of digital face images (\mathbf{X}_{Gal}). Here, \mathbf{X}_{Gal} contains the gallery images arranged in a column wise manner, such that x_{Gal}^i corresponds to the i^{th} gallery image. Feature extraction is performed for all the gallery images using the transform matrix \mathbf{T} :

$$\mathbf{A}_{Gal} = \mathbf{T} \times \mathbf{X}_{Gal} s.t. \|\mathbf{A}_{Gal}\|_0 \leq \tau \quad (7)$$

Similarly, the representation of a given test image (x_{Skull}) is calculated as follows:

$$a_{Skull} = \mathbf{T} \times x_{Skull} s.t. \|a_{Skull}\|_0 \leq \tau \quad (8)$$

This is followed by calculating the Euclidean distance between the probe feature and the gallery features, in order to obtain the identity having the least distance:

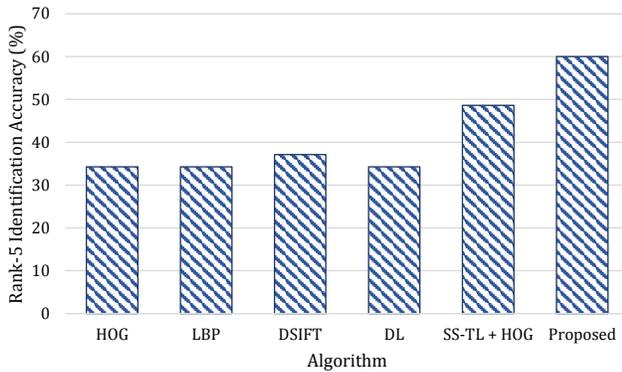
$$\arg \min_i \|a_{Skull} - a_{gal}^i\|_2^2 \quad (9)$$

The identity of x_{gal}^i is returned as the identity of the given probe.

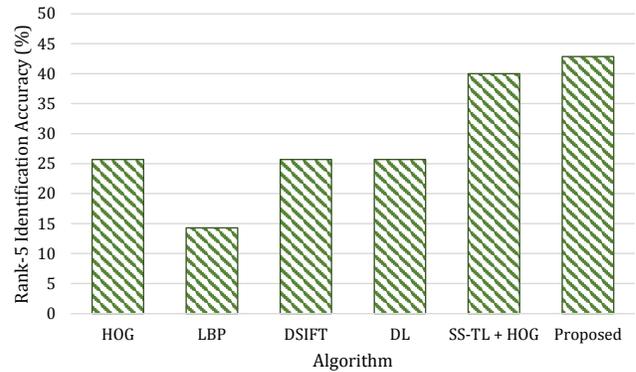
3. Experiments and Results

There exists only one publicly available dataset for performing skull to digital face image matching, the IdentifyMe dataset [8]. The dataset contains 464 skull images, inclusive of 35 mated skull images and their corresponding face images. Figure 5 presents sample images of the IdentifyMe dataset. The authors have also provided two protocols and baseline results for evaluating the performance of any algorithm. The two protocols are as follows:

- **Protocol-1:** This protocol evaluates the performance of a given algorithm on the 35 mated pairs of skull and digital face image. The results are reported with five fold cross validation.



(a) Protocol-1



(b) Protocol-2

Figure 6: Bar graph representing the rank-5 identification accuracy for both the protocols on IdentifyMe dataset. SS-TL+HOG corresponds to the current state-of-the-art approach [8].

- **Protocol-2 (extended gallery):** This protocol attempts to mimic a real world scenario where the gallery consists of more subjects as compared to those present in the dataset. An extended gallery of 993 subjects is used for evaluation, such that for a given fold, the gallery contains 1000 subjects (993 extended and 7 from the IdentifyMe dataset) at test time. Similar to the previous protocol, five fold cross validation is performed for evaluation.

The proposed Shared Transform Model is evaluated on both the protocols. Face images of IdentifyMe dataset are detected using Viola Jones face detector [17], followed by geometric alignment with the skull images. Data augmentation is performed by combinations of flipping across the y-axis and modifying the brightness or contrast of images. Features are extracted from the pre-processed images, which are then used as input for the proposed Shared Transform Model. The learned model is then used with the test images, along with Euclidean distance for performing matching with the final transformed features (as shown in Section 3.1).

Baseline results with Histogram of Oriented Gradients (HOG) [3], Local Binary Pattern (LBP) [11], Dense Scale Invariant Feature Transform (DSIFT) [2], and Dictionary Learning (DL) [7] have been reported on the two protocols of IdentifyMe dataset. Since the best baseline performance is obtained with Histogram of Oriented Gradients [3] features, the proposed Shared Transform Model is applied on the same. Table 1 and Figure 6 present the rank-1 and rank-5 identification accuracy (%) for both the protocols. The key results are as follows:

- Table 1 presents the rank-1 identification accuracy (%) of the proposed model on the IdentifyMe dataset for both the protocols, along with the comparative algo-

Table 1: Rank-1 identification accuracy (%) on the IdentifyMe dataset. Baseline results have directly been taken from [8].

Algorithm		Protocol-1	Protocol-2
Baseline	LBP [11]	28.5	14.2
	DSIFT [2]	28.5	20.0
	DL [7]	31.4	22.9
	HOG [3]	34.2	25.7
	US-TL + HOG [8]	37.1	34.2
	SS-TL + HOG [8]	42.9	37.1
Proposed + HOG		51.4	42.9

rithms (averaged across five folds). It can be observed that with HOG features as input, the proposed Shared Transform Model achieves an accuracy of 51.4% and 42.9% on the two protocols, respectively.

- Shared Transform Model with HOG features showcases an improvement of almost 17% as compared to the identification accuracy obtained with HOG features only. This demonstrates that the proposed model is able to encode discriminative features, useful for classification. Figure 7 presents sample skull and digital face image pairs correctly classified by the proposed model for Protocol-1. It is interesting to observe that the proposed Shared Transform Learning algorithm is able to model the large intra-class variations between data of different domains.
- As compared to the current state-of-the-art (SS-TL + HOG) [8], the proposed model presents an improvement of around 9% and over 5% for the first and the second protocol, respectively. An improvement of around 15% is observed as compared to US-TL + HOG

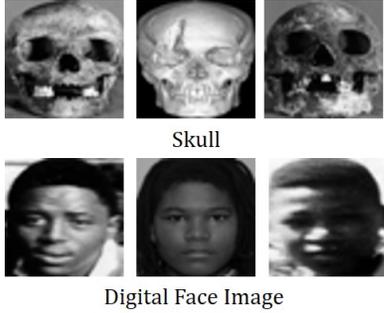


Figure 7: Sample skull and digital face images correctly identified by the proposed Shared Transform Model. It can be observed that the proposed model is able to encode the heterogeneous variations present across both domains.

as well.

- The focus of the proposed model is to learn a single transform matrix, whereas SS-TL learns two transforms and a classifier. The proposed model thus requires learning lesser number of parameters, which in turn results in improved performance for the given problem having limited training data. For an input of dimension $n \times 1$, the proposed Shared Transform Model requires learning $n \times n$ parameters (shared transform matrix) for transforming the input to a dimension of $n \times 1$. On the other hand, the SS-TL model requires learning $n \times n$ parameters for the skull transform, $n \times n$ for the digital face images transform, and $n \times 1$ parameters for the perceptron based classifier. Therefore, the number of parameters learned in the proposed model is less than half as compared to SS-TL [8]. We believe that given the small sample size problem of skull recognition, this reduction in the number of parameters improves the learning capability and generalizability of the model.
- Figure 6 presents the rank-5 identification accuracy (%) of the proposed Shared Transform Model with HOG features on the IdentifyMe dataset. Accuracies pertaining to both the protocols are reported, where, the proposed Shared Transform Model outperforms other algorithms by reporting an improvement of at least 3% as compared to the state-of-the-art results.
- Figure 8 presents sample weights learned by the proposed Shared Transform Model for the task of skull recognition, on raw pixel values. It is observed that the model encodes variations observed in different skull and digital face images. For instance, one can observe different overall structures, and variations specifically in the jaw and eye region of the weights. This demonstrates that the proposed model learns variables encompassing the variations observed in the input data.

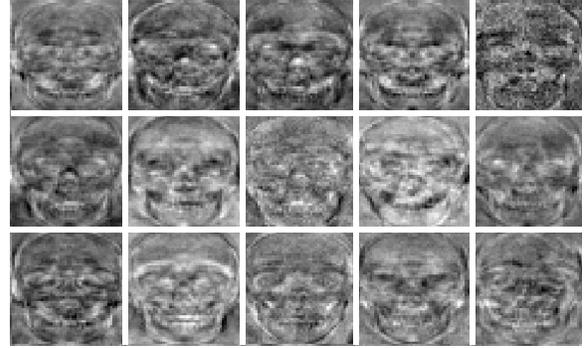


Figure 8: Sample weights learned by the Shared Transform Model on raw pixels. It demonstrates that the model learns variations with respect to the structure of the input data.

- Upon analyzing the standard deviation of the algorithms across all folds, we observe that the proposed Shared Transform Model with HOG features presents a standard deviation of 7.1% on Protocol-1. On the other hand, identification with HOG features results in a standard deviation of 11.9%. For Protocol-2, the proposed model achieves a standard deviation of 6.3%, whereas identification using HOG features results in a standard deviation of 23.4%. The reduced standard deviation obtained across all folds for both the protocols showcases the robustness of the proposed approach for performing skull to digital face image matching.

4. Conclusion

The task of skull to digital image matching is of vital importance for law enforcement in missing children cases, unsolved cold cases, or in instances of mass casualties. In all these real-world scenarios, the proposed algorithm can be valuable since it matches a skull image with gallery face images, if they exist. This can potentially reduce the manual processing burden on forensic artists and forensic anthropologists. The heterogeneous nature of the problem, along with the availability of limited training data, makes the problem challenging. In this research, a novel Shared Transform Model is proposed for performing skull to digital face image matching. The proposed model learns a “shared transform” for skulls and digital face images, while reducing the intra-class variations between the learned features. Experimental evaluation on the publicly available IdentifyMe dataset showcases the efficacy of the proposed model by achieving improved performance for the two protocols given with the dataset.

5. Acknowledgment

This research is partially supported by Infosys Center for Artificial Intelligence at IIIT-Delhi. S. Nagpal is partially supported through TCS PhD Fellowship.

References

- [1] W. Aulsebrook, M. Íscan, J. Slabbert, and P. Becker. Superimposition and reconstruction in forensic facial identification: a survey. *Forensic Science International*, 75(2):101–120, 1995.
- [2] A. Bosch, A. Zisserman, and X. Muñoz. Image classification using random forests and ferns. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.
- [3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 886–893, 2005.
- [4] S. Damas, O. Cordón, O. Ibáñez, J. Santamaría, I. Alemán, M. Botella, and F. Navarro. Forensic identification by computer-aided craniofacial superimposition: A survey. *ACM Computing Surveys*, 43(4):27:1–27:27, 2011.
- [5] F. Duan, Y. Yang, Y. Li, Y. Tian, K. Lu, Z. Wu, and M. Zhou. Skull identification via correlation measure between skull and face shape. *IEEE Transactions on Information Forensics and Security*, 9(8):1322–1332, 2014.
- [6] M. I. Huete, O. Ibez, C. Wilkinson, and T. Kahana. Past, present, and future of craniofacial superimposition: Literature and international surveys. *Legal Medicine*, 17(4):267–278, 2015.
- [7] D. D. Lee and H. S. Seung. Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:788–791, 1999.
- [8] S. Nagpal, M. Singh, A. Jain, R. Singh, M. Vatsa, and A. Noore. On matching skulls to digital face images: A preliminary approach. In *IEEE International Joint Conference on Biometrics*, pages 813–819, 2017.
- [9] S. Nagpal, M. Singh, R. Singh, M. Vatsa, A. Noore, and A. Majumdar. Face sketch matching via coupled deep transform learning. In *IEEE International Conference on Computer Vision*, pages 5429–5438, 2017.
- [10] S. Nagpal, M. Vatsa, and R. Singh. Sketch recognition: What lies ahead? *Image and Vision Computing*, 55:9–13, 2016.
- [11] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [12] S. Ouyang, T. Hospedales, Y.-Z. Song, X. Li, C. C. Loy, and X. Wang. A survey on heterogeneous face recognition: Sketch, infra-red, 3D and low-resolution. *Image and Vision Computing*, 56:28–48, 2016.
- [13] S. Ravishankar and Y. Bresler. Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging. *SIAM Journal on Imaging Sciences*, 8(4):2519–2557, 2015.
- [14] S. Ravishankar, B. Wen, and Y. Bresler. Online Sparsifying Transform Learning-Part I: Algorithms. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):625–636, 2015.
- [15] P. Reavy and L. Egan. Skull remains unidentified 3 years after its discovery. <https://www.ksl.com/?nid=148&sid=44886383>. Posted: 3rd July, 2017.
- [16] P. Tu, R. Book, X. Liu, N. Krahnstoeber, C. Adrian, and P. Williams. Automatic face recognition from skeletal remains. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2007.
- [17] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.