On Matching Skulls to Digital Face Images: A Preliminary Approach

Shruti Nagpal[1], Maneet Singh[1], Arushi Jain[1], Richa Singh[1,2], Mayank Vatsa[1,2], Afzel Noore[2]
[1]IIIT Delhi, India, [2]West Virginia University
{shrutin, maneets, arushi13023, rsingh, mayank}@iiitd.ac.in, afzel.noore@mail.wvu.edu

Abstract

Forensic application of automatically matching skull with face images is an important research area linking biometrics with practical applications in forensics. It is an opportunity for biometrics and face recognition researchers to help the law enforcement and forensic experts in giving an identity to unidentified human skulls. It is an extremely challenging problem which is further exacerbated due to lack of any publicly available database related to this problem. This is the first research in this direction with a two-fold contribution: (i) introducing the first of its kind skull-face image pair database, IdentifyMe, and (ii) presenting a preliminary approach using the proposed semi-supervised formulation of transform learning. The experimental results and comparison with existing algorithms showcase the challenging nature of the problem. We assert that the availability of the database will inspire researchers to build sophisticated skull-to-face matching algorithms.

1. Introduction

"Some say, 'That's such an old case, why do it?' I say, 'Why not?' They still deserve their names back." - Joe Mullins, National Center for Missing and Exploited Children.

In 2012, law enforcement authorities found the remains of a young girl in Opelika, Oklahoma[1]. From the skull of the girl, forensic experts reconstructed the face (top left in Figure 1) and released it to the public. After few years, they also created the computerized composite of the girl and shared it with the public. However, more than five years have passed and the identity of the girl is still unknown.

This case and several other such cases show that identification is not only required by the living people but it is also important for the dead. Identification for the living has been well studied and multiple biometric modalities provide efficient and accurate methods for identification. However, dead bodies and remains of individuals who have suffered unfortunate death due to crime or natural calamities, often do not have the luxury of having biometrics modalities intact. Fingerprints [4] and iris [12] may be decomposed while dental and DNA patterns may not exist in medical or national identification databases. Further, there are several cases where only some body parts are available, for instance head, hand or leg. While the bones in hand or leg can be used to determine DNA and, to a certain extent eth-

Figure 1: Clay reconstruction and 3D facial composite of a girl whose remains were found in Opelika, Oklahoma.

Figure 2: Skeleton, reconstructed clay composite, and actual face image of two individuals. The images are taken from the Internet.

[1]https://identifyus.org/en/cases/9834

nicity and gender as well, however, forensic experts can reconstruct a lot more information if the skull is available. They can determine the DNA, age, gender (depending on the age), and reconstruct the entire face which may be very useful in forensic and law enforcement investigation.

This research paper explores how the biometrics and face recognition community can help forensic experts in identifying the person from the skull. Given the skull of a person, forensic experts first determine the age, gender and ethnicity of the person using several morphological measures [14], followed by generating the facial reconstruction from the skull using clay sculpture. This process is generally performed manually; however, recently some of the labs have started using 3D reconstruction software such as FaceIT and ReFace [3]. Figure 2 shows the skeleton, clay reconstruction, and original face images (from left to right) from two solved cases[2] [3]. This process, when performed manually, takes a little over a week for one skull while generating computerized 3D reconstruction takes 3-4 days. Given this process, face matching can be performed in two ways: (i) matching the skull with face images, and (ii) matching the composite with face images. Matching reconstructed composite with face images might appear easier, however it requires an expert forensic anthropologist and at least 3-4 days for generating the composite.

In this research, we examine the feasibility of matching skull to face images, that may be available in a missing person database. We postulate that matching skull to face images can be considered as another dimension/covariate in heterogeneous face recognition research with forensic applications. The first and foremost requirement of pursuing this research is a database of skull and corresponding face images. Therefore, we first prepared the *IdentifyMe - Skull and Face Image Database* which contains mated pairs of skull and face images from 35 individuals and 429 independent skull images. The database will be released to the research community to promote further research in this area. We next propose a preliminary algorithm using transform learning to match skull with face images, along with establishing the baseline skull-to-face matching performance of several existing algorithms. The results show that the problem at hand is a challenging yet feasible research problem, requiring creation of a larger database and specifically designed algorithms to correctly determine the identity of the person with high accuracy/precision.

2. Proposed IdentifyMe Dataset

While existing research has focused on creating facial reconstructions from skulls, no work has been done to automate the matching of skull images to digital face images. This is the first work which aims to match a given skull image to a large number of face images to retrieve the identity of the given skull image. Due to the lack of an existing dataset for the above problem, we have prepared the proposed *IdentifyMe* dataset. The proposed dataset contains 464 skull images, and is divided into two parts: (i) skull and digital face image pairs, and (ii) unlabeled supplementary skull images. The first component consists of pairs collected from real world examples of solved skull identification cases, while the second part contains unlabeled skull images. Details of the two components of the proposed IdentifyMe dataset are provided below:

2.1. Skull and Face Image Pairs

A total of 35 skull images and their corresponding face images are collected from various sources. Some of these pairs correspond to real world cases where a skull was found and later identified to belong to a missing person. Owing to the unavailability of such kind of real world data, this component also consists of some famous reconstructed face and skull image pairs. Figure 3(a) presents some sample skull-digital face image pairs from the proposed IdentifyMe dataset. Unavailability of high quality, well-illuminated digital face images for real world cases further renders the given problem challenging.

2.2. Unlabeled Supplementary Skull Images

The second component of the proposed dataset consists of 429 unlabeled, supplementary skull images. While it is difficult to obtain skull-digital image pairs, however, one can also obtain unlabeled skull images from the Internet. Although the identities of the skull images of this component are unknown, these set of images can be utilized for unsupervised learning based algorithms. Owing to the heterogeneity of the problem, and large variations observed between the domains of digital face images and skulls, this component of unlabeled supplementary skull images can thus be used to reduce the domain gap between the two.

The proposed IdentifyMe dataset will be made publicly available to the research community[4]. Moreover, this dataset will also be made available as a *contributory* dataset, wherein researchers are invited to add to the existing dataset by providing some of the images (obtained from the Internet or other sources). We will first ensure that the images are not already available in the database and then release the new set of images as part of the database with acknowledgment to the contributors. This will ensure incremental growth in the resources and foster research in the domain of skull to digital face image matching.

[2] http://catyanaskoryfalsetti.com/forensic-art/

[3] https://archives.fbi.gov/archives/about-us/lab/forensic-science-communications/fsc/jan2001/phillips.htm

[4] http://iab-rubric.org/resources/identifyme.html

(a) The first row corresponds to the digital face images, while the second contains their respective skull images.

(b) Unlabeled supplementary skull images.

Figure 3: Sample images from the proposed IdentifyMe - Skull and Face Image Database.

2.3. Protocols

In order to standardize further research on skull to digital photo matching, two protocols are provided for the proposed dataset. In real world scenarios of missing person, it is unlikely for law enforcement agencies to perform face verification (1:1 matching) with skull images. Therefore, two identification protocols are defined on the proposed dataset. Each protocol consists of five-fold cross validation in order to remove any bias of a particular train-test partition. As is the case in real world, the digital photos correspond to the gallery images (or the database), and the skull images correspond to the probes. The two protocols are defined as follows:

Protocol-1: The first protocol utilizes only the 35 skull-face image pairs. These pairs are divided into five folds for performing five fold cross validation. Each fold consists of seven pairs, such that four folds are used for training, and the remaining one fold is used for testing.

Protocol-2: In real world scenarios, a skull image is required to be identified against a large database of face images from different gender, age and ethnicity. Therefore, in order to mimic this real world scenario, the second protocol consists of an extended gallery, where 993 additional digital images are added in the gallery (to obtain a gallery of 1000 images). Therefore, each fold in this protocol consists of 1000 gallery images (digital face images) and 7 probe images (skull images).

The first protocol attempts to perform identification within the proposed dataset only, while the second protocol mimics the real world scenario of having a much larger dataset against which a given input will be matched. The above defined train-test partitions, for all five folds will be provided along with the dataset to the research community.

3. Proposed Algorithm

For skull to face matching, there are two major restrictions in building automated algorithms: (i) limited training samples per subject and (ii) overall limited training data. Due to the sensitive nature of the problem, getting multiple training samples from each subject is exceptionally challenging. The problem is further exacerbated with availability of limited labeled training data (specifically, no training data before this research). Therefore, data intensive algorithms such as deep learning will be extremely demanding and may overfit. On the other hand, algorithms which do not require training such as handcrafted features or relatively less training such as dictionary learning or transform learning can be explored. In this research, we propose transform learning based skull to face matching by utilizing the domain specific knowledge (i.e. face) and problem specific knowledge (i.e. skull).

Ravishankar and Bresler [9] proposed transform learning, an analysis equivalent of dictionary learning. It is a representation learning technique to learn a transformation space, $\mathbf{T}$ which is used to produce a representation, $\mathbf{Z}$ for given input samples, $\mathbf{X}$. It is mathematically written as:

$$\begin{aligned} \underset{\mathbf{T},\mathbf{Z}}{\arg\min} \left\|\mathbf{TX}-\mathbf{Z}\right\|_F^2 + \lambda \left(\epsilon \left\|\mathbf{T}\right\|_F^2 - log\det \mathbf{T}\right) \\ s.t. \left\|\mathbf{Z}\right\|_0 \leq \tau \end{aligned} \tag{1}$$

A penalty term, in the form of ℓ_2-norm regularization of the transformation matrix, $\mathbf{T}$ is added in order to balance the scale. The $(log\det \mathbf{T})$ term is the log-determinant regularizer [6], added to prevent possible degenerate solutions. Existing papers [10, 11] show the following alternating minimization approach to learn $\mathbf{T}$ and $\mathbf{Z}$:

1. Initialize $\mathbf{T}$, $\mathbf{Z}$
2. Fix $\mathbf{T}$, learn $\mathbf{Z}$ as:

$$\mathbf{Z} \leftarrow \underset{\mathbf{Z}}{\arg\min} \left\|\mathbf{TX}-\mathbf{Z}\right\|_F^2, \; such\; that \; \left\|\mathbf{Z}\right\|_0 \leq \tau \tag{2}$$

3. Fix $\mathbf{Z}$, learn $\mathbf{T}$ as:

$$\mathbf{T} \leftarrow \underset{\mathbf{T}}{\arg\min} \left\|\mathbf{TX}-\mathbf{Z}\right\|_F^2 + \lambda \left(\epsilon \left\|\mathbf{T}\right\|_F^2 - log\det \mathbf{T}\right) \tag{3}$$

Natively, transform learning is an unsupervised representation learning approach. In this research, we first extend the formulation and propose Semi-Supervised Transform Learning. We next propose two algorithms for skull to face

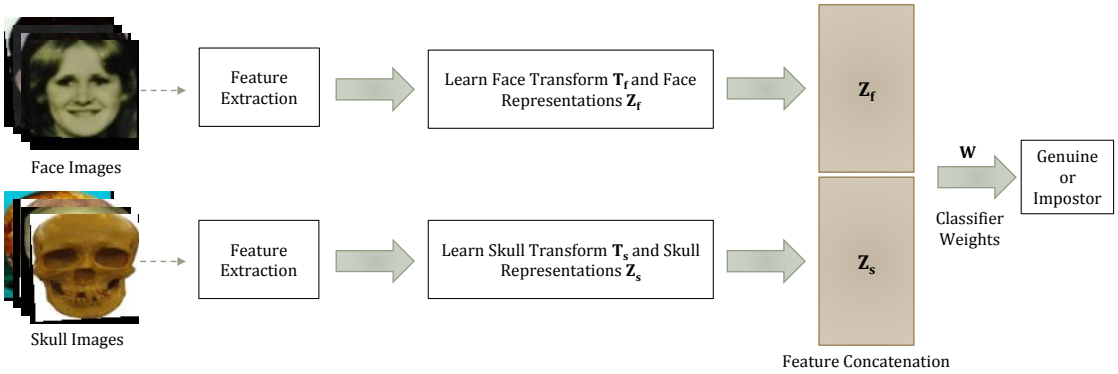

Figure 4: Illustrating the steps involved in the proposed Supervised Transform Learning for skull to face image recognition. Input images or features of skulls and face images are used to learn transform matrices ($\mathbf{T_s}, \mathbf{T_f}$) and corresponding representations ($\mathbf{Z_s}, \mathbf{Z_f}$). The learned representations are then used to learn the classifier weights $\mathbf{W}$.

identification: (a) using Unsupervised Transform Learning (US-TL) and (b) using Semi-Supervised Transform Learning (SS-TL).

3.1. Skull to Face Matching using Unsupervised Transform Learning (US-TL)

The proposed unsupervised transform learning based algorithm utilizes the traditional transform learning model to learn representations for the face images and skull images independently. The proposed algorithm is explained as follows:

1. Given input training face images $\mathbf{X_f}$, a transformation matrix, $\mathbf{T_f}$ and representation $\mathbf{Z_f}$ are learned.

$$\underset{\mathbf{T_f},\,\mathbf{Z_f}}{\arg\min} \|\mathbf{T_f X_f} - \mathbf{Z_f}\|_F^2 + \lambda(\epsilon \|\mathbf{T_f}\|_F^2 - log\det \mathbf{T_f}) \tag{4}$$

   Equation 4 is learned using unlabeled face images with the aim of learning effective representations in a transformed space.

2. Similarly, a transform learning model is trained for skull images $\mathbf{X_s}$ to learn representations $\mathbf{Z_s}$ and transformation matrix, $\mathbf{T_s}$:

$$\underset{\mathbf{T_s},\,\mathbf{Z_s}}{\arg\min} \|\mathbf{T_s X_s} - \mathbf{Z_s}\|_F^2 + \lambda(\epsilon \|\mathbf{T_s}\|_F^2 - log\det \mathbf{T_s}) \tag{5}$$

   In order to learn $\mathbf{T_s}$ and $\mathbf{Z_s}$, unlabeled skull images are used (details are given in implementation details).

3. Euclidean distances of the final representations, $\mathbf{Z_f}$ and $\mathbf{Z_s}$ are used to perform identification or match skull images to face images.

3.2. Skull to Face Matching using Semi-Supervised Transform Learning (SS-TL)

The unsupervised transform learning approach does not specifically encode "matchability" and "supervision" with respect to domain specific knowledge (i.e. face) and problem specific knowledge (i.e. skull). We extend the unsupervised transform learning formulation to semi-supervised transform learning for matching a given skull image to faces. Figure 4 illustrates the overview of the proposed algorithm.

In this research, the unsupervised transform learning is extended using a perceptron like classifier-weight learning. For face and skull input data ($\mathbf{X_f}$ and $\mathbf{X_s}$), the modified formulation is expressed as:

$$\begin{aligned} \underset{\mathbf{T_s},\,\mathbf{Z_s},\,\mathbf{T_f},\,\mathbf{Z_f},\,\mathbf{W}}{\arg\min} & \|\mathbf{T_f X_f} - \mathbf{Z_f}\|_F^2 \\ + \lambda(\epsilon \|\mathbf{T_f}\|_F^2 - log\det \mathbf{T_f}) & + \|\mathbf{T_s X_s} - \mathbf{Z_s}\|_F^2 \\ + \lambda(\epsilon \|\mathbf{T_s}\|_F^2 - log\det \mathbf{T_s}) & + \mathbf{W}[\mathbf{Z_f Z_s}] \end{aligned} \tag{6}$$

where, $\mathbf{W}[\mathbf{Z_f Z_s}]$ is the supervision term which enables supervision by encoding class information to perform improved matching of skull to face representations, and $\|\mathbf{T_f X_f} - \mathbf{Z_f}\|_F^2$ and $\|\mathbf{T_s X_s} - \mathbf{Z_s}\|_F^2$ terms are used to learn representations. Since supervised training samples are very limited, we further extend the training process which utilizes both unsupervised and supervised samples as follows:

1. The first step in the proposed algorithm involves learning two independent representations and transforms ($\mathbf{Z_f}$, $\mathbf{Z_s}$, $\mathbf{T_f}$, and $\mathbf{T_s}$) using Equations 4 and 5 for face and skull, respectively. This is performed using unlabeled face and skull data.

2. In Equation 6, initialize $\mathbf{T_f}, \mathbf{Z_f}$ and $\mathbf{T_s}, \mathbf{Z_s}$ using the values obtained from unlabeled training data obtained in Step 1. Initialize $\mathbf{W}$ with ones.

3. Using labeled training data, recompute $\mathbf{T_s}, \mathbf{Z_s}, \mathbf{T_f}, \mathbf{Z_f}$ and compute $\mathbf{W}$.

In the proposed semi-supervised training, we utilize both domain specific and problem specific knowledge and later incorporate supervision for improved performance. To train the model, alternating minimization approach is utilized.

3.3. Implementation Details

Unsupervised Transform Learning: Face images from CMU Multi-PIE [5] dataset have been used as unlabeled face images to train Equation 4 and learn $\mathbf{T_f}$ and $\mathbf{Z_f}$. The same images are also used for pre-training the model in Equation 5. Unlabeled Supplementary Skull images from the proposed *IdentifyMe* dataset are used to perform unsupervised training to learn $\mathbf{T_s}$ and $\mathbf{Z_s}$.

Semi-Supervised Transform Learning: $\mathbf{T_f}$, $\mathbf{Z_f}$ and $\mathbf{T_s}$, $\mathbf{Z_s}$ are learned using Equation 4 and 5 as explained above in an unsupervised manner for face and skull images respectively. To incorporate supervision and learn the weight matrix, $\mathbf{W}$, the training partition of the skull and face image pairs from the proposed *IdentifyMe* dataset is utilized.

Detection and Registration: Digital faces are detected using Haar Cascade approach [13] whereas skulls are manually detected. Area-based alignment is performed using the Image Alignment Toolbox[5], in order to register the skull and face images. In the proposed algorithms, we have used pixel values for $\mathbf{X_f}$ and $\mathbf{X_s}$ (represented as US-TL+Pixels) and SS-TL + Pixels), and HOG features (represented as US-TL + HOG and SS-TL + HOG). Experiments are performed on a workstation with Xeon 2.7GHz processor and 64GB RAM under MATLAB environment.

4. Experiments and Results

Experiments are performed on the above protocols defined in Section 2.3 to present baseline results on the proposed dataset, along with the performance of the proposed algorithm. Identification results are shown with the following hand-crafted features and learned representations: (i) Histogram of Oriented Gradients (HOG) [2], (ii) Local Binary Pattern (LBP) [8], (iii) Dense Scale Invariant Feature Transform (DSIFT) [1], and (iv) Dictionary Learning [7]. In Dictionary Learning, the dictionary is trained on a subset of CMU Multi-PIE [5] face dataset, containing frontal well-illuminated face images.

Once the features are extracted using the above algorithms, Euclidean distance is utilized for performing identification. For the proposed transform learning algorithm, results are presented with both Unsupervised Transform Learning (US-TL) and Semi-Supervised Transform Learning (SS-TL). Since the proposed models can extract features from any input data, results are shown with raw pixels (images as it is) and HOG features as input. Data augmentation is performed on the gallery images by flipping across the y-axis and varying the brightness.

[5]http://iatool.net/

Table 1: Rank-1 identification accuracies (%) obtained for the two protocols on the proposed IdentifyMe dataset.

Algorithm	Protocol-1	Protocol-2
HOG [2]	34.2	25.7
LBP [8]	28.5	14.2
DSIFT [1]	28.5	20.0
DL [7]	31.4	22.9
Proposed Transform Learning		
US-TL + Pixels	31.4	25.7
SS-TL + Pixels	34.2	28.6
US-TL + HOG	37.1	34.2
SS-TL + HOG	**42.9**	**37.1**

Figure 5 presents bar graphs summarizing the rank-1 and rank-5 identification accuracies and Table 1 tabulates the rank-1 identification accuracies for the two protocols on the proposed dataset. Average accuracies across the five folds are reported. It is observed that:

- For the first protocol, where the gallery consists of only seven face images corresponding to the seven skull probes, the maximum rank-1 identification accuracy obtained, without using the proposed algorithm is 34.2% (HOG). Figure 5(a) presents the corresponding bar graph of accuracies with all the features.

- The proposed Semi-Supervised Transform Learning (SS-TL) with HOG features as input presents the best performance by achieving a rank-1 identification accuracy of 42.9% (Table 1). The improvement of more than 8% over the accuracy achieved by HOG features can directly be attributed to the proposed formulation.

- Table 1 also presents the performance of Unsupervised Transform Learning (US-TL) with pixel values and HOG features as input. For both cases, SS-TL performs at least 3% better than US-TL. This motivates the inclusion of supervision in the proposed model for performing skull to digital image matching.

- Figure 6 presents the score distribution of the genuine and impostor pairs using HOG features for the first protocol, obtained across all five folds. No clear distinction can be observed between the scores of the two classes, further motivating the challenging nature of the problem at hand. While the number of genuine scores having large value and impostor scores having low value of Euclidean distance is less, however, the distribution in the middle range is overlapping.

- Figure 5(b) presents the bar graph obtained for the second protocol, where an extended gallery of 993 subjects

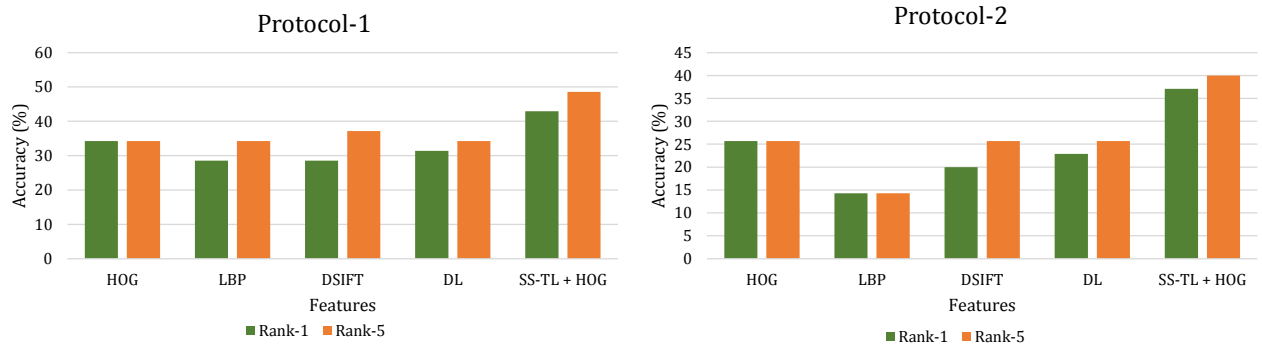

Figure 5: Bar graphs displaying the identification accuracies (%) obtained at rank-1 and rank-5 for the two protocols.

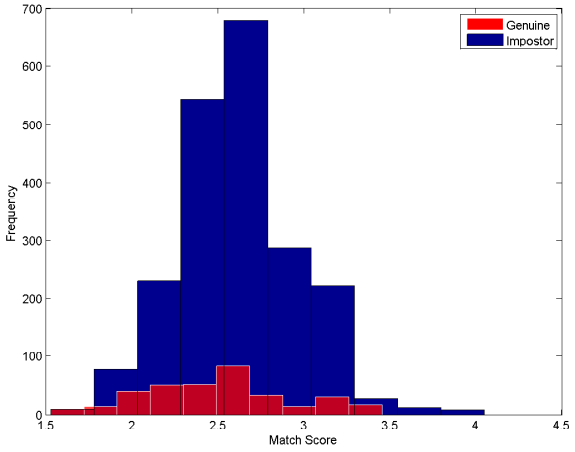

Figure 6: Score distribution of the Euclidean distance values obtained for the genuine and impostor face-skull pairs using HOG features for the first protocol.

is also used (i.e, total gallery size is 1000). Similar to the previous protocol, Histogram of Oriented Gradients (HOG) obtains the highest identification accuracy of 25.7%. Using HOG features as input, the proposed SS-TL model achieves a rank-1 identification accuracy of 37.1%. The proposed model displays an improvement of more than 10%, as compared to HOG features based classification.

- The improved performance of the transform learning models obtained at rank-1 and rank-5 (Figure 5) further motivates learning based algorithms utilizing the supplementary unlabeled component of the proposed IdentifyMe dataset. Since the information content in skulls and digital face images varies significantly, learning from the supplementary unlabeled skull images helps in reducing these variations.

- The drop in the best performing rank-1 accuracy from 42.9% (protocol-1) to 37.1% (protocol-2) motivates the challenging nature of the problem. In real world cases, where the law enforcement agencies have a large dataset of images against which a given probe image needs to be matched, improved algorithms are required to address this challenging problem.

5. Conclusion

Determining the identity of human remains, particularly skull, is a long standing problem in forensic applications. Forensic experts first create a facial reconstruction from the skull using clay technique or computerized 3D methods. The composite can then be shared with the media and matched across already available face databases including the missing person database. This research attempts to build an algorithm for matching skull to digital face images. A novel database, *IdentifyMe* database, has been prepared by the authors which consists of more than 450 skull images, including 35 real world skull-face image pairs. Two protocols emulating the real world scenarios, along with their baseline results have also been reported to facilitate research in this direction. Due to the difference in information content of skull and digital images, a Semi-Supervised Transform Learning (SS-TL) algorithm has been proposed for performing identification of skull images. For both the protocols, the proposed SS-TL algorithm achieves improved performance, in comparison with other baseline results. In order to encourage further research in this area, the IdentifyMe dataset will be made publicly available to the research community as a *contributory* dataset, wherein, other researchers would be encouraged to contribute and use the dataset.

6. Acknowledgment

S. Nagpal is partially supported through TCS PhD fellowship. M. Vatsa and R. Singh are partially supported through Infosys Center for Artificial Intelligence.

References

[1] 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.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005.

[3] S. Decker, J. Ford, S. Davy-Jow, P. Faraut, W. Neville, and D. Hilbelink. Who is this person? A comparison study of current three-dimensional facial approximation methods. *Forensic Science International*, 229(1):161.e1 – 161.e8, 2013.

[4] A. W. Fox. Post mortem fingerprinting and unidentified human remains. *Journal of Forensic and Legal Medicine*, 24:46 – 47, 2014.

[5] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. *Image and Vision Computing*, 28(5):807–813, 2010.

[6] B. Kulis, M. Sustik, and I. Dhillon. Learning low-rank kernel matrices. In *International Conference on Machine Learning*, pages 505–512, 2006.

[7] D. D. Lee and H. S. Seung. Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:788–791, 1999.

[8] 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.

[9] S. Ravishankar and Y. Bresler. Learning sparsifying transforms. *IEEE Transactions on Signal Processing*, 61(5):1072–1086, 2013.

[10] 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.

[11] S. Ravishankar and Y. Bresler. Online Sparsifying Transform Learning-Part II: Convergence Analysis. *IEEE Journal of Selected Topics in Signal Processing*, 9(4):637–646, 2015.

[12] M. Trokielewicz, A. Czajka, and P. Maciejewicz. Human iris recognition in post-mortem subjects: Study and database. In *IEEE International Conference on Biometrics Theory, Applications and Systems*, pages 1–6, 2016.

[13] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[14] B. A. Williams and T. L. Rogers. Evaluating the accuracy and precision of cranial morphological traits for sex determination. *Journal of Forensic Sciences*, 51(4):729–735, 2006.