Heterogeneity Aware Deep Embedding for Mobile Periocular Recognition

Rishabh Garg[1][*], Yashasvi Baweja[1][*], Soumyadeep Ghosh[1], Richa Singh[1], Mayank Vatsa[1], Nalini Ratha[2]
[1]IIIT-Delhi, India, [2]IBM TJ Watson, USA

[1]{rishabh15076, yashasvi15116, soumyadeepg, rsingh, mayank}@iiitd.ac.in, [2]ratha@us.ibm.com

Abstract

Mobile biometric approaches provide the convenience of secure authentication with an omnipresent technology. However, this brings an additional challenge of recognizing biometric patterns in an unconstrained environment including variations in mobile camera sensors, illumination conditions, and capture distance. To address the heterogeneous challenge, this research presents a novel heterogeneity aware loss function within a deep learning framework. The effectiveness of the proposed loss function is evaluated for periocular biometrics using the CSIP, IMP and VISOB mobile periocular databases. The results show that the proposed algorithm yields state-of-the-art results in a heterogeneous environment and improves generalizability for cross-database experiments.

1. Introduction

Mobile devices are ubiquitous and they are used for various applications such as mobile banking, e-business and social media. These devices store confidential and critical data which if lost/stolen can cause harm to the user. Therefore, secure, convenient and fast authentication methods are required to unlock the devices. Most of the modern mobile devices rely on biometric based authentication [17] such as face and fingerprint recognition to validate the identity of the user. However, biometric authentication on mobile devices pose several challenges. A primary challenge in acquiring the biometric data from mobile phones is that it is highly unconstrained. For touch-less sensing (e.g. capturing faces), the quality of the image can be adversely affected by factors such as variation in illumination conditions, distance from the subject, indoor/outdoor scenarios, quality of the front and back camera, and motion blur due to movement of the device/subject. Different mobile sensors for capturing biometric data pose a cross sensor matching problem, as different camera sensors have different imaging properties. This introduces heterogeneity in the captured data (e.g., indoor vs outdoor, front camera vs back camera

---

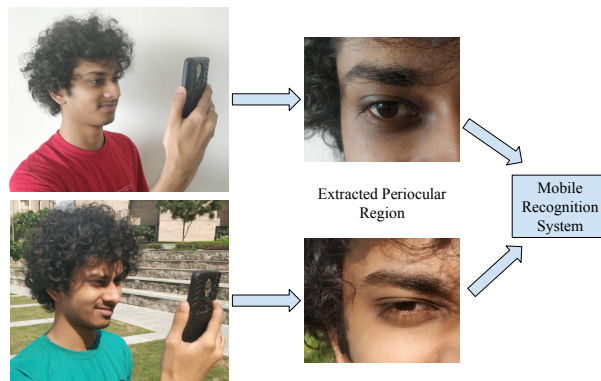*Equal contributions by Garg and Baweja

Figure 1: Data captured from mobile devices in indoor and outdoor conditions result in large variations.

resolution), and it makes biometric recognition on mobile devices an interesting and challenging problem.

Periocular region as a biometric modality [5, 21] has been gaining attention. It refers to using the regions around the eye for identity recognition. The periocular region is generally available even in unconstrained scenarios with a non cooperative subject and it can be especially useful in situations where the other information such as face is partially occluded. Figure 1 illustrates the use of mobile periocular recognition in unconstrained environments. It requires no additional capturing overhead which is useful while capturing using a mobile device. Feasibility of periocular region as a biometric trait was explored by Park *et al*. [21]. Thereafter, there has been significant research advancements in this area. Detailed surveys of periocular recognition are provided by Alonso-Fernandez *et al*. [2] and Nigam *et al*. [20]. A large number of techniques have performed periocular recognition on data obtained with high quality sensors in constrained conditions but there has been increasing focus on the less constrained scenarios as well. Many popular methods relied on hand crafted features like HOG, SIFT and LBP for the periocular and iris information [5, 20]. Tan *et al*. [32] use filters applied on input data for providing discriminative features for segmentation and recognition. Nie *et al*. [19] use convolutional restricted Boltzmann machine

along with handcrafted feature extraction for improved performance.

Recently, deep convolutional neural networks have gained immense popularity for ocular recognition. Zhao and Kumar [36] use explicit semantic information to extract better features and improve performance of the CNN. Proença *et al*. [22] generate artificial samples belonging to multiple classes by interchanging ocular parts from different subjects for data augmentation thereby improving the training process. Several works have also explored the problem of periocular recognition by capturing data using mobile devices. De Frietas *et al*. [7] model the inter session variability in the data from the enrollment time to the test time. Raghavendra *et al*. [23] utilize coupled autoencoders and Maximum Response (MR) based texture features for mobile periocular recognition. Another approach by Raja *et al*. [24] used pooling of sparse filtered features. Zhang *et al*. [35] use the fusion of iris and periocular region information with weighted concatenation to obtain a joint representation.

In this paper, a novel heterogeneity aware deep embedding framework for periocular recognition is proposed specifically for scenarios where the images are captured in unconstrained settings. The proposed method works by obtaining the heterogeneity invariant feature representations of the periocular images via a deep convolutional neural network. The deep CNN model is trained via the proposed heterogeneous aware loss metric based on the identity of the subjects and tries to enforce a margin between the clusters of images of a particular identity/class in the embedding space. The embeddings of the same subject/classes are brought close to each other and that of other subjects are pushed away from each other in the output embedding space of the deep CNN model. In addition to that, the loss function ensures that the model produces heterogeneity aware embeddings. Experiments are performed on three popular periocular databases and comparison with existing algorithms demonstrate state-of-the-art results. The remaining paper is arranged as follows: Section 2 contains details of the proposed algorithm. The database used and experiment protocols are discussed in Section 3 while the results are discussed in Section 4.

2. Proposed Algorithm

In mobile periocular recognition, heterogeneity may occur due to illumination variations, change in subject to camera distances, and sensor variations. In this section, we illustrate a novel periocular recognition algorithm which trains a deep convolutional neural network model using the proposed heterogeneity aware loss metric. This results in a highly discriminative model producing heterogeneity aware embeddings suitable for matching periocular images captured in unconstrained scenarios. Figure 2 illustrates the steps involved in the proposed pipeline.

2.1. Motivation

In the homogeneous/ideal scenarios, the vanilla Triplet Loss [28] can be used which enforces a margin α on the embeddings for a given set of three images known as a triplet. Let (I_{a1}, I_{a2}, I_{b1}) be a triplet where I_{a1} is the anchor image of identity/class 'a', I_{a2} is the positive image which belongs to the same person (identity/class 'a') and I_{b1} is a negative sample of identity/class 'b'. Let $g(I_x)$ be the feature embeddings of image I_x and τ is the set of all triplets and $[z]_+$ is $max(0, z)$. The Triplet loss [28] aims to minimize the following:

$$\sum_{\forall T \in \tau} \left[\|g(I_{a1}) - g(I_{a2})\|_2^2 - \|g(I_{a1}) - g(I_{b1})\|_2^2 + \alpha \right]_+$$

(1)

$$\forall (I_{a1}, I_{a2}, I_{b1}) \in \tau$$

For a model to produce heterogeneity aware embeddings, it should learn to discriminate between images of different identities as well as bring closer the embeddings of similar identities even in the presence of domain variation at the image level. Such a model should not work with just a single negative sample in the particular triplet. Instead, if the model learns to differentiate between an image of 'a' and every image of 'b' (here 'a' and 'b' are two different identities) then the model generalizes better because it has to enforce a margin with all the embeddings of the negative class as opposed to a single negative sample.

In order to represent all the embeddings of the negative class, mean embedding of the negative class can be incorporated in the vanilla triplet loss. This means that essentially the centroid of the cluster of images of a negative class is separated from the positive class images. The loss function for the same is as follows:

$$L =$$

$$\left[\|g(I_{a1}) - g(I_{a2})\|_2^2 - \left\| g(I_{a1}) - \frac{\sum_{i=1}^k g(I_{bi})}{k} \right\|_2^2 + \alpha_1 \right]_+$$

(2)

where, I_{a1}, I_{a2} belong to class 'a' and I_{bi} is the ith image of class 'b' (I_{a1} serves as the anchor), $\frac{\sum_{i=1}^k g(I_{bi})}{k}$ represents the mean of all the embeddings of a random negative identity 'b'.

2.2. Heterogeneity aware embedding space

Equation 2 only incorporates mean embeddings in the same domain and there is no factor of domain/covariate variations. In order to incorporate domain/covariate variation in the model, images needs to be added from different domains for both identities 'a' and 'b'.
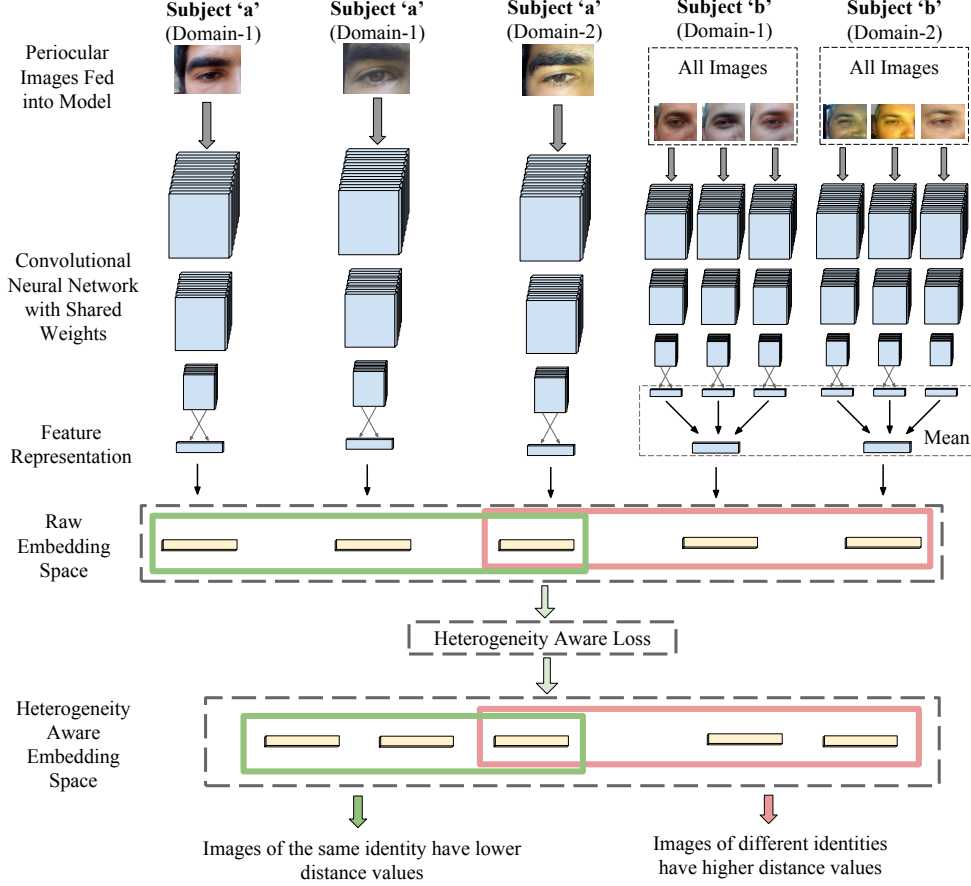
Figure 2: Training the proposed model: periocular images pertaining to different identities are utilized to forward pass through the deep CNN model with shared weights. During training, the loss function (Figure 3) optimizes the feature representations so that the ones of the same identity are close to each other (i.e. reduce intra-class variations) while others are pushed further apart in the output embedding space of the deep CNN model. a and b refer to different subjects and domain 1 and 2 refer to different image capture scenarios such as indoor/outdoor and with flash/without flash.

Let p and q be the factors of domain variation which we want to incorporate together in the model. Equation 2 with the covariate can be expressed as:

$$L_1 = \left[\|g(I_{a1}^p) - g(I_{a2}^p)\|_2^2 - \left\| g(I_{a1}^p) - \frac{\sum_{i=1}^k g(I_{bi}^p)}{k} \right\|_2^2 + \alpha_1 \right]_+ \tag{3}$$

For multiple domains, we would still like to minimize the distance between the embeddings for same identities and increase it for different identities. This implies, minimize $\|g(I_{a1}^p) - g(I_{a3}^q)\|_2^2$ and maximize $\|g(I_{a1}^p) - g(I_{b1}^q)\|_2^2$ where I_x^q is an image in different domain and $g(I_x^q)$ is its respective deep CNN model embedding. This means that the cluster of embeddings of a particular class is essentially shrunk as the embeddings are brought closer while the centroid of the cluster of a negative class is pushed away in the embedding space. Hence, the loss equation to train a domain invariant

representation can be expressed as:

$$L_2 = \left[\|g(I_{a1}^p) - g(I_{a3}^q)\|_2^2 - \|g(I_{a1}^p) - g(I_{bi}^q)\|_2^2 + \alpha_2 \right]_+ \tag{4}$$

Representing the negative class by the mean embedding, Equation 4 can be expressed as:

$$L_2 = \left[\|g(I_{a1}^p) - g(I_{a3}^q)\|_2^2 - \left\| g(I_{a1}^p) - \frac{\sum_{i=0}^k g(I_{bi}^q)}{k} \right\|_2^2 + \alpha_2 \right]_+ \tag{5}$$

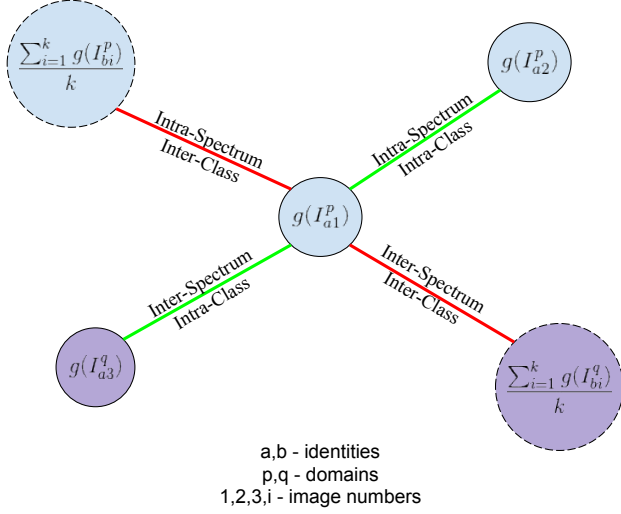The final loss equation for creating heterogeneity aware embedding space would be ($\mathcal{L} = L_1 + L_2$):

a,b - identities
p,q - domains
1,2,3,i - image numbers

Figure 3: Illustrating the proposed heterogeneity aware loss

$$\mathcal{L} =$$
$$\left[\|g(I_{a1}^p) - g(I_{a2}^p)\|_2^2 - \left\|g(I_{a1}^p) - \frac{\sum_{i=1}^k g(I_{bi}^p)}{k}\right\|_2^2 + \alpha_1\right]_+$$
$$+$$
$$\left[\|g(I_{a1}^p) - g(I_{a3}^q)\|_2^2 - \left\|g(I_{a1}^p) - \frac{\sum_{i=0}^k g(I_{bi}^q)}{k}\right\|_2^2 + \alpha_2\right]_+$$
$$(6)$$

This loss function can be used to train a domain invariant representation in a deep CNN model, which can be utilized to train for both homogeneous (same domain) and heterogeneous (cross-domain) scenarios.

2.3. Implementation Details

The CNN architecture used for training is LightCNN29 [34]. The network consists of 29 convolutional layers with 3×3 filters. There are 4 pooling layers and the feature representation (embedding) layer is 256 dimensional. The optimization of the gradient of the loss function is performed via Adam optimizer [14] at a learning rate of $1e^{-5}$ which is slowly decayed. The values of both the summations in the loss are clipped to have a lower bound of 0. The data to be provided to the CNN is sampled randomly from the data available for training and composed into the required tuple. For the experiments, both α_1 and α_2 have been set to 0.4.

3. Databases and Experimental Protocols

The efficacy of our model is evaluated on two datasets for unconstrained heterogeneous data captured from mobile devices: the CSIP database [27] and the VISOB database [26]. Additionally, we have reported results on the IIITD Multi-spectral Periocular Database [29] which has data in different spectrums collected using different sensors including a handheld nightvision camera to show the effectiveness of the proposed algorithm in a general heterogeneous data acquisition scenario. Figure 4 shows sample images from these databases.

3.1. CSIP Database

The Cross-sensor iris and periocular dataset [27] contains images captured from 4 different mobile phones- Sony Ericsson Xperia Arc, Apple iPhone 4, ThL W200 and Huawei U8510. Images taken from each sensor (mobile phone camera) is further divided into categories denoting front/rear camera and flash/no flash. The dataset has 2004 right periocular images pertaining to 50 different subjects. For this dataset, we carry out two experiments, cross sensor and cross illumination periocular recognition. For cross sensor tasks, we train the algorithm on one-vs-all setup, where all images from Apple iPhone 4 serve as one domain, and all images from the remaining sensors are considered as second domain. Training and testing partition is done such that images of subjects 1-40 are used for training and images of subjects from 41-50 form the testing set. Additionally we test the proposed algorithm on cross-illumination tasks, such that all the images in the presence of flash form one domain and images captured without flash correspond to different domain. Train test split is similar according to the above protocol. Results for both the experiments are reported in Tables 1 and 2.

3.2. VISOB Dataset

The VISOB database [26] is a large scale dataset from the VISOB ICIP2016 Challenge. It consists of images from 550 subjects captured via the front facing camera of 3 different devices - iPhone 5s, Samsung Note 4 and Oppo N1 in 3 different illumination conditions namely, regular office light, dim light and natural daylight settings. The data was collected in two visits. Only Visit 1 is publicly available. It contains a total of 48,250 images as a part of the enrollment set and 46,797 images as a part of the verification set across all devices and conditions. We perform two experiments on the dataset. **(a):** In the first experiment, for training, all the images in the enrollment set are used and for testing, the images in the verification set act as probes for the enrolled images via which identification is performed similar to [1]. **(b):** In order to compare with [37], the training and testing was performed only on the images captured via the iPhone in day light conditions (as per the protocol used in [37]).

3.3. IIITD Multi-spectral Periocular Database

The IIITD IMP dataset [29] has images captured in three spectrums - visible, near-infrared and night vision, making

Figure 4: Sample images from the CSIP [27], VISOB [26], and IMP [29] datasets.

a total of 1220 images. With 62 subjects in each spectrum and 5 different images corresponding to each subject, the dataset contains 310 images each in the visible and the NIR spectrum. Resolution of the visible spectrum images is 601×301 and the NIR images are of 540×260 each. To demonstrate the effectiveness of the proposed approach, no training is performed on this database. The proposed algorithm is evaluated by using the model trained on cropped images of the CASIA NIR-VIS 2.0 face database [16]. This is done in order to keep the protocols consistent (to perform comparison) with other cross-spectral periocular recognition methods namely Behera et al [4] and Ramaiah et al [25].

4. Experimental Results

The proposed model is evaluated on the datasets discussed in Section 3, and compared with other state-of-the-art algorithms. For CSIP[1] dataset, the performance of the proposed algorithm is compared with Triplet Loss [28] trained in the same way described in Section 3. The training protocol is exactly consistent with the one used for the proposed algorithm. For the cross-illumination and cross-sensor experiments (Table 1 and Table 2) the proposed algorithm achieves a Rank 1 Accuracy of 87.33% and 89.53%, respectively. It outperforms [28] by over 10% and 5%, respectively. This illustrates the superiority of the method in generating embeddings that are invariant to the large heterogeneity in the data. Furthermore, apart from the deep learning methods, we also show the comparison with hand-crafted features such as Histogram of Oriented Gradients (HOG) [6] and Daisy features (similar to SIFT) [33]. The

---

[1]Kandaswamy *et al*. [12] has reported results on this database, but the protocol used in their work is transfer learning based. Santos et al. [27] had performed cross-sensor experiments, but evaluated their algorithm on the entire dataset. Since the proposed method requires training, a direct comparison with [27] is not feasible. Monteiro *et al*. [18] have also computed the results on this dataset, however cross sensor experiments were not performed

Table 1: Results on the CSIP dataset for cross-sensor mobile periocular recognition tasks.

Algorithm	Identification	Verification GAR@f FAR	
	Rank-1(%)	f=0.1%	f=10%
HOG [6]	62.79	2.85	27.84
DAISY [33]	62.40	2.49	33.57
Schroff *et al.* [28]	84.10	12.87	65.64
Proposed	**89.53**	**18.23**	**75.15**

Table 2: Results on the CSIP dataset for cross-illumination mobile periocular recognition tasks.

Algorithm	Identification	Verification GAR@f FAR	
	Rank-1(%)	f=0.1%	f=10%
HOG [6]	73.85	3.19	27.21
DAISY [33]	57.26	3.42	29.80
Schroff *et al.* [28]	77.42	10.17	59.66
Proposed	**87.33**	**14.53**	**83.19**

results presented in Tables 1 and 2 corroborate the effectiveness of the proposed model.

Table 3 summarizes the Rank 1 accuracies of the proposed method on the VISOB Database [26] for the experiment **(a)** (described in section 3.2). The proposed method outperforms the current state-of-the-art for all devices and lighting conditions, significantly. Table 4 summarizes the results obtained on the VISOB database for experiment **(b)**. For comparison with Zhao *et al.* [37] the same experimental protocol is followed and the results obtained are reported on the same fold. The proposed method obtained an improvement of over 10% over the state-of-the-art EER.

The results of the IMP dataset are summarized in Table 5. It is important to note that no training is performed on this dataset and the reported results are used to illustrate the effectiveness of the model to generate embeddings which can match identities irrespective of the heterogeneity. The method achieves a Genuine Accept Rate of 82.97% at 10% False Accept Rate. As shown in Table 5, the proposed approach outperforms the state-of-the-art by a very large margin. Results are also compared with the deep learning technique [28] and the proposed method achieves rank 1 accuracy of 61.2% as compared to 49.36% obtained by [28].

Apart from the accuracies observed, we have made following observations:

Cross-Database Performance: In order to compare the performance of the proposed approach with state-of-the-art algorithms [4, 25] for the IMP dataset, we performed testing on this dataset without training on any image of this dataset. The deep CNN model was trained on the CASIA NIR-VIS 2.0 dataset [16]. Periocular images were extracted from the face images of this dataset for training. This train-

(a) IMP Database	(b) CSIP Database (cross illumination)

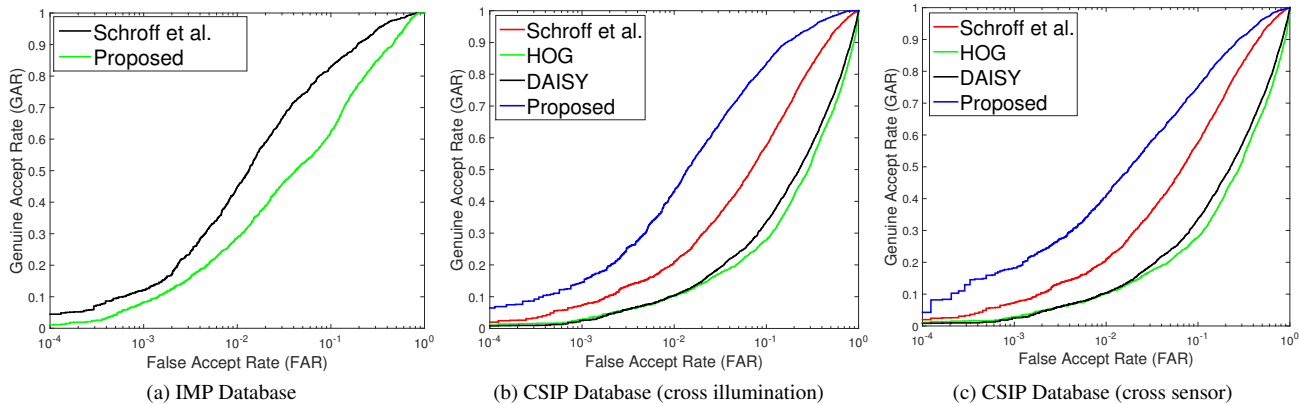(c) CSIP Database (cross sensor)

Figure 5: ROC curves showing verification accuracies on IMP and CSIP databases

Table 3: Rank 1 accuracy on the VISOB Database for experiment 1 with all images.

Phone	Condition	Rank 1 Accuracy (%)			
		Ahuja [1]	Proposed	Ahuja [1]	Proposed
		Left		Right	
Samsung	Office	90.45	**94.30**	91.53	**94.71**
	Day	92.44	**97.15**	92.97	**98.47**
	Dim	93.12	**97.19**	93.61	**98.04**
iPhone	Office	93.54	**94.97**	93.89	**95.88**
	Day	95.98	**96.36**	94.82	**96.06**
	Dim	96.09	**96.69**	96.14	**96.54**
Oppo	Office	90.79	**91.55**	90.23	**90.75**
	Day	94.21	**97.66**	94.81	**97.25**
	Dim	96.31	**97.28**	96.15	**97.07**

Table 4: Results on the VISOB Database with iPhone in daylight

Algorithm	Rank 1 Accuracy(%)	EER (%)
Texton [32]	-	4.80
PPDM [31]	-	5.03
SCNN [36]	-	3.30
Zhao *et al.* [37]	-	1.47
Proposed	**99.41**	**1.32**

Table 5: Results on the IMP dataset for cross-spectrum periocular recognition tasks.

Algorithm	Identification	Verification GAR@f FAR	
	Rank-1(%)	f=0.1%	f=10%
Ramaiah *et al.* [25]	-	-	18.35
Behara *et al.* [4]	-	-	25.03
Schroff *et al.* [28]	49.36	8.23	62.27
Proposed	**61.20**	**12.07**	**82.97**

ing was performed with spectrum as the heterogeneity and then the trained model was utilized for testing on the entire IMP dataset. This mimics a cross-database train-test scenario. As shown in Table 5, the proposed algorithm pro-

duces state-of-the-art results, which shows that our model is generalizable to datasets on which no fine-tuning or training is performed. It should also be noted that the CASIA and IMP datasets contain subjects pertaining to different ethnicities and the images are collected using different sensors. High verification performance with cross-database testing is a strong indication of the generalizability of the algorithm.

Hard Mining: Most deep metric learning algorithms [11, 28] are heavily dependent on hard mining of samples for training. However, the proposed method, produces better results than one of the most popular deep metric learning algorithms [28] without any hard-mining. This saves a huge amount of training time and is a testament to the efficacy of the proposed algorithm.

Testing Time: On Intel Core $i7$ workstation with 32GB of RAM and NVIDIA GTX 1080ti GPU, the average time for matching a pair of images is 50.5 microseconds.

5. Conclusion and Future Research

Mobile periocular recognition requires addressing heterogeneity due to illumination variations, subject-to-camera distances, sensor variations, and indoor-outdoor variations. To address this research challenge, a heterogeneity aware loss is proposed to train deep CNN model which helps in creating domain invariant embedding space. The proposed algorithm for periocular recognition in unconstrained environments achieves state-of-the-art results. Although the results are shown on periocular recognition tasks, the proposed loss metric can also be extended for other recognition tasks such as recognizing faces with disguise variations [9, 15, 30], heterogeneous face recognition [8, 10], and iris/periocular recognition with multiple cameras or covariates [3, 13].

6. Acknowledgement

References

[1] K. Ahuja, R. Islam, F. A. Barbhuiya, and K. Dey. Convolutional neural networks for ocular smartphone-based biometrics. *PRL*, 91:17–26, 2017.

[2] F. Alonso-Fernandez and J. Bigun. A survey on periocular biometrics research. *PRL*, 82:92–105, 2016.

[3] S. S. Arora, M. Vatsa, R. Singh, and A. Jain. On iris camera interoperability. In *IEEE BTAS*, pages 346–352, 2012.

[4] S. S. Behera, M. Gour, V. Kanhangad, and N. Puhan. Periocular recognition in cross-spectral scenario. In *IEEE IJCB*, pages 681–687, 2017.

[5] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh. Periocular biometrics: When iris recognition fails. In *IEEE BTAS*, pages 1–6, 2010.

[6] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE CVPR*, volume 1, pages 886–893, 2005.

[7] T. de Freitas Pereira and S. Marcel. Periocular biometrics in mobile environment. In *IEEE BTAS*, pages 1–7, 2015.

[8] T. I. Dhamecha, P. Sharma, R. Singh, and M. Vatsa. On effectiveness of histogram of oriented gradient features for visible to near infrared face matching. In *IAPR ICPR*, pages 1788–1793, 2014.

[9] T. I. Dhamecha, R. Singh, M. Vatsa, and A. Kumar. Recognizing disguised faces: Human and machine evaluation. *PLoS one*, 9(7):e99212, 2014.

[10] S. Ghosh, T. I. Dhamecha, R. Keshari, R. Singh, and M. Vatsa. Feature and keypoint selection for visible to near-infrared face matching. In *IEEE BTAS*, pages 1–7, 2015.

[11] A. Hermans, L. Beyer, and B. Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

[12] C. Kandaswamy, J. C. Monteiro, L. M. Silva, and J. S. Cardoso. Multi-source deep transfer learning for cross-sensor biometrics. *Neural Computing and Applications*, 28(9):2461–2475, 2017.

[13] R. Keshari, S. Ghosh, A. Agarwal, R. Singh, and M. Vatsa. Mobile periocular matching with pre-post cataract surgery. In *IEEE ICIP*, pages 3116–3120, 2016.

[14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] V. Kushwaha, M. Singh, R. Singh, M. Vatsa, N. Ratha, and R. Chellappa. Disguised faces in the wild. In *IEEE CVPRW*, volume 8, 2018.

[16] S. Li, D. Yi, Z. Lei, and S. Liao. The CASIA NIR-VIS 2.0 face database. In *IEEE CVPRW*, pages 348–353, 2013.

[17] N. Maček, I. Franc, M. Bogdanoski, and A. Mirković. Multimodal biometric authentication in IoT: Single camera case study. In *BISEC*, 2016.

[18] J. C. Monteiro, R. Esteves, G. Santos, P. T. Fiadeiro, J. Lobo, and J. S. Cardoso. A comparative analysis of two approaches to periocular recognition in mobile scenarios. In *ISVC*, pages 268–280, 2015.

[19] L. Nie, A. Kumar, and S. Zhan. Periocular recognition using unsupervised convolutional rbm feature learning. In *IAPR ICPR*, pages 399–404, 2014.

[20] I. Nigam, M. Vatsa, and R. Singh. Ocular biometrics: A survey of modalities and fusion approaches. *Information Fusion*, 26:1–35, 2015.

[21] U. Park, A. Ross, and A. K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *IEEE BTAS*, pages 1–6, 2009.

[22] H. Proença and J. C. Neves. Deep-prwis: Periocular recognition without the iris and sclera using deep learning frameworks. *IEEE TIFS*, 13(4):888–896, 2018.

[23] R. Raghavendra and C. Busch. Learning deeply coupled autoencoders for smartphone based robust periocular verification. In *IEEE ICIP*, pages 325–329, 2016.

[24] K. B. Raja, R. Raghavendra, and C. Busch. Collaborative representation of deep sparse filtered features for robust verification of smartphone periocular images. In *IEEE ICIP*, pages 330–334, 2016.

[25] N. P. Ramaiah and A. Kumar. On matching cross-spectral periocular images for accurate biometrics identification. In *IEEE BTAS*, pages 1–6, 2016.

[26] A. Rattani, R. Derakhshani, S. K. Saripalle, and V. Gottemukkula. In *IEEE ICIP*, 2016.

[27] G. Santos, E. Grancho, M. V. Bernardo, and P. T. Fiadeiro. Fusing iris and periocular information for cross-sensor recognition. *PRL*, 57:52–59, 2015.

[28] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE CVPR*, pages 815–823, 2015.

[29] A. Sharma, S. Verma, M. Vatsa, and R. Singh. On cross spectral periocular recognition. In *IEEE ICIP*, pages 5007–5011, 2014.

[30] R. Singh, M. Vatsa, and A. Noore. Face recognition with disguise and single gallery images. *Elsevier IVC*, 27(3):245–257, 2009.

[31] J. M. Smereka, V. N. Boddeti, and B. V. Kumar. Probabilistic deformation models for challenging periocular image verification. *IEEE TIFS*, 10(9):1875–1890, 2015.

[32] C.-W. Tan and A. Kumar. Towards online iris and periocular recognition under relaxed imaging constraints. *IEEE TIP*, 22(10):3751–3765, 2013.

[33] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE TIFS*, 32(5):815–830, 2010.

[34] X. Wu, R. He, Z. Sun, and T. Tan. A light cnn for deep face representation with noisy labels. *IEEE TIFS*, 13(11):2884–2896, 2018.

[35] Q. Zhang, H. Li, Z. Sun, and T. Tan. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE TIFS*, 13(11):2897–2912, 2018.

[36] Z. Zhao and A. Kumar. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE TIFS*, 12(5):1017–1030, 2017.

[37] Z. Zhao and A. Kumar. Improving periocular recognition by explicit attention to critical regions in deep neural network. *IEEE TIFS*, 13(12):2937–2952, 2018.