A Secure Visual-thermal Fused Face Recognition System Based on Non-Linear Hashing

Xingbo Dong, KokSheik Wong, Zhe Jin
School of Information Technology
Monash University Malaysia
Sunway, Malaysia
{xingbo.dong, wong.koksheik, jin.zhe}@monash.edu

Jean-Luc Dugelay
Department of Digital Security
EURECOM
Sophia-Antipolis, France
jld@eurecom.fr

***Abstract*—In this paper, we propose a secure visual-thermal fused face recognition system using non-linear hashing. To extract features from both thermal and visible facial images, a deep neural network model pre-trained by visible images, namely InsightFace, is utilized in extracting deep features from both thermal and visible images. Next, we investigate into the effectiveness of using nonlinear hashing in protecting deep features extracted from both thermal and visible face images. To further boost the accuracy performance of the facial recognition system under unfavorable environment, feature- and score-level fusion of thermal and visible images for face matching are studied. The performance of different application scenarios are tested on the EURECOM VIS-TH face dataset. Experiment results suggest that: 1) feature- and score-level fusion techniques are effective in achieving higher accuracy under unfavorable situation; 2) non-linear hashing offers additional layer of protection, namely, privacy preservation, to face image. We also found that the deep model trained by using visible images is applicable to thermal images for feature extraction, which is particularly useful because there is no large thermal dataset available to train deep neural network.**

***Index Terms*—Face recognition, thermal face, fusion, biometric template protection**

I. Introduction

Face recognition (FR) has been a long-standing interest for identity management since the first publication in Nature [1] in year 1888. Recent trend shows that FR has huge potential beyond identity management, such as human computer interaction and entertainment. This is attributed mainly to the acquisition process of face biometrics, which is easy and non-intrusive, while other biometric modalities such as fingerprint cannot always be captured, e.g., heavy labor worker or born without fingerprint (i.e., adermatoglyphia). Furthermore, face biometrics is compliant with human minds and do not require much cooperation from the user in order to function.

FR technology evolves from handcraft method such as eigenface [2] and local binary pattern [3] to deep learning approach [4]. Currently, deep learning approach outperforms handcraft methods in most FR contests [5]. However, the accuracy performance can still be affected by many factors, which can generally be divided into the environmental and user factors. Specifically, environmental factor includes light changes and unfavorable illumination, while user factor includes emotions, occlusions and head pose variation. Another key challenges regarding the FR deep model are the issues of privacy and security. In [6], a neighborly de-convolutional neural network (NbNet) is proposed to reconstruct the face images from deep features. It is reported that face image reconstructed by NbNet can be easily matched with the original counterpart, where the true accept rate (TAR) can achieve 95.20% when comparing the reconstructed images against the original face images, which is a serious threat.

Cancelable biometrics is one of the approaches for biometric template protection. It is developed to address the security and privacy risks when a biometric template is compromised. It refers to the irreversible transformation that can alter the biometric templates for ensuring security and privacy. A cancelable biometrics scheme should satisfy four requirements, namely, non-invertibility, revocability, non-linkability, and performance preservation [7].

To address the aforementioned environmental issues, the combination of visible and thermal information has been demonstrated to increase face recognition rates, particularly in the presence of adverse conditions [8]. We also believe that the combination of visible and thermal image will be a promising technique in future FR system, e.g. cross-spectral forensics, mobile face identification. Here, we also show that dual acquisition can achieve better accuracy in the design of a cancelable FR system when compared to traditional approaches based on a single modality (visible or thermal). Specifically, facial features are extracted from both visible and thermal images using the public software *InsightFace* [9]. The features are then encrypted by NMDSH, which is a recently proposed non-invertible transformation function within the context of concealable face templates [10].

Our work makes the following contributions:

- We propose a framework of visible and thermal fused FR system which can overcome the drawbacks of using visible face under limited or unfavorable illumination condition. Both feature- and score-level fusion techniques

This research is supported by EU Horizon 2020 - Marie Sklodowska-Curie Actions through the project entitled Computer Vision Enabled Multimedia Forensics and People Identification (Project No. 690907, Acronym: IDENTITY), Fundamental Research Grant Scheme (FRGS/1/2018/ICT02/MUSM/03/3). We also gratefully acknowledge the support of NVIDIA Corporation for its GPU grant.

978-1-7281-1817-8/19/$31.00 2019 IEEE

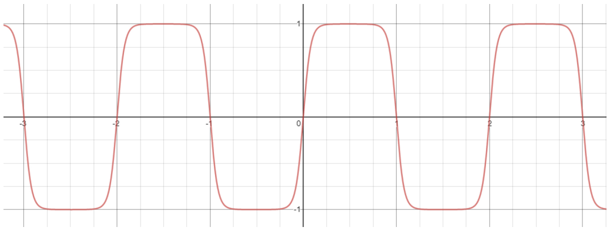


Fig. 1: Visualizing non-linear function $q(x)$.

are effective in achieving higher accuracy under unfavorable situation.

- Non-linear hashing is used in the system to achieve privacy preservation. The face features are protected against privacy invasion.
- Both visible and thermal face images can be processed by same deep learning model pre-trained by using visible face images. The findings in this study could be beneficial toward a secure FR system with high accuracy performance.

II. Non-linear hashing

The deployment of various FR applications leads to the creation of many databases of faces. Once compromised, these databases will lead to security and privacy issues. Therefore, biometric template protection technology is developed to address these issues. In this paper, we investigate into the effectiveness of non-linear hashing in protecting deep features generated from both thermal and visible face images. The non-linear hashing considered here is based on multi-dimensional spectral hashing (MDSH) for face template protection proposed in [10]. Specifically, MDSH is extended from the graph-based hamming embedding [11], [12] for cancelable biometrics.

In our current work, two realizations in [10], namely, slim multi-dimensional spectral hashing (SMDSH), and non-linear multi-dimensional spectral hashing (NMDSH) are studied to protect deep face features.

In MDSH scheme, assume that x_i and x_j are two data point in Euclidean space. A hashing function should generate their corresponding hash code y_i and y_j in Hamming space by approximating the Hamming distance between y_i and y_j to the Euclidean distance of x_i and x_j. The Euclidean distance between data points x_i and x_j is represented as affinity matrix $W(i,j) = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}}$. In addition, each bit should be independently and uniformly distributed. Based on this assumption, MDSH can be taken as a binary matrix factorization problem of the affinity matrix, and it can be solved by the spectral relaxation technique, hence becomes a eigenvector problem [12].

The SMDSH algorithm generally consists of three steps as the original MDSH does:

1) Compute the single-dimension eigen functions $\phi_{ij}(x(i))$ and the corresponding eigen value λ_{ij} on the training dataset using:
$$\phi_{ij}(x(i)) = \sin\left(\frac{\pi}{2} + \frac{j\pi}{b_i - a_i} x(i)\right), \tag{1}$$
and
$$\lambda_{ij} = \exp\left(-\frac{\delta^2}{2}\left|\frac{j\pi}{b_i - a_i}\right|^2\right). \tag{2}$$
where $\phi_{ij}(x(i))$ is the j-th eigen function of the i-th coordinate, and λ_{ij} is the corresponding eigenvalue.
2) Sort λ_{ij} in ascending order, and select the top k indices to form the set $A = \{(i_1, i_1), (i_2, i_2), \cdots, (i_k, i_k)\}$.
3) Encode each data point x in the test dataset using $y_{ij}(x) = \sin(\phi_{ij}(x))$ for all $(i,j) \in A$.

Although MDSH maintains the accuracy of the deep face feature, it is vulnerable to similarity-based attack (SA) due to its distance preserving property [13]. Therefore, inspired by [13], a *softmod* activation layer is added to SMDSH to achieve a nonlinear MDSH (hereinafter referred to as 'NMDSH') as follows:
$$y = q(\phi_{ij}(x)), \tag{3}$$
and
$$q(x) = \frac{2}{1 + e^{-8\sin(\alpha\pi\ x)}} - 2, \tag{4}$$
where $q(x)$ is a nonlinear *softmod* activation function (see Fig. 1) and the nonlinear rate α is an empirical parameter defined by user. In a nutshell, both SMDSH and NMDSH employ one-way function to transform the original features into some binary forms, hence achieving the template protection goal. In addition, the generated binary vectors can also be used for indexing or searching due to efficient computation of Hamming distance.

III. Thermal Imagery

Visible face image is largely sensitive to illumination. This leads to the deterioration of recognition accuracy when operating under unfavorable illumination. Therefore, most FR systems can only achieve high recognition accuracy under constrained environment [14]. On the other hand, infrared spectrum (IR) imagery is known to be robust against illumination effect. It can capture image under both day and night conditions, hence it is widely regarded as a plausible option in case of unfavorable illumination.

Generally, the IR spectrum can be divided into active IR band and passive (thermal) IR band according to the response of various detectors (see Fig. 2) [15]. The active band (0.7-2.5μm) consists of near infrared (NIR) and the short wave infrared (SWIR) spectrum, while thermal IR band includes Mid-Wave (MWIR, 3-5μm) and the Long-Wave infrared (LWIR, 7-14μm) bands. MWIR captures reflective and emissive properties of the face skin while LWIR primarily captures the emitted radiation or heat energy. Therefore, LWIR is insensitive to variation in illumination.

However, datasets containing thermal LWIR face images are rare. EURECOM VIS-TH Face (VIS-TH) [8] is a recently

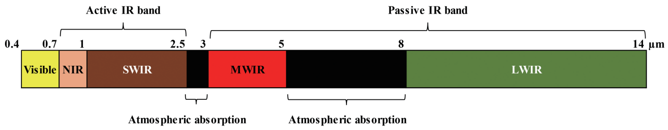


Fig. 2: Electromagnetic spectrum map.

TABLE I: Different variation in VIS-TH

Category	Variations
Expression	Neutral (N), Happy (EH), Angry (EA), Sad (ES), Surprised (ESp)
Action	Eyes Closed (AEC), Open Mouth (AOM)
Pose	Up (PU), Down (PD), Left (PL), Right (PR)
Occlusion	Optical Glasses (OOG), Sunglasses (OSG), Hat (OH), Hand on Mouth (OHM), Hand on Eye (OHE)
Light	Light Up (LLU), Light Right (LLR), Light Left (LLL),Dark (LD), Room light (LR)

proposed face dataset which uses a newly developed dual sensor, namely FLIR®Duo™ R (2017), to simultaneously capture face images and videos in both visible and thermal LWIR spectra. This guarantees that face image of visible and thermal spectra can be captured from the same person (i.e. synchronized). A total of 2100 images under various settings (see Table I) were collected via FLIR®Duo™ R from 50 subjects of different age, gender and ethnicity.

IV. Fusion of Visible and Thermal Images

It is known that biometric fusion can lead to performance gain [16]. An evidence of fusion for thermal and visible face images has been reported in [8], showing the improvement of rank-1 recognition accuracy. However, the performance evaluation of the fusion strategies have not been investigated on the secure FR system (e.g., FR with biometric template protection). Here, FR with only visible or thermal image is called single modality FR, while FR with both visible and thermal images is called multi-modalities. Typically, a single modality cancelable FR system consists of sensor, feature extractor, non-invertible parameterized transformation function and the matcher. Fig. 3(a) shows a typical single modality cancelable FR system, where F is a one-way transformation function (e.g., SMDSH and NMDSH). Note that the visible and thermal face images are used independently to implement the single modality scheme. On the other hand, for the multi-modalities cancelable FR, three schemes (one cross-spectral, two fusion schemes) are designed:

1) The hashed code generated from the visible face features are enrolled in the gallery while the hashed code from thermal face is used as a probe;
2) The fused feature is formed by concatenating the visible and thermal face features, which is then hashed and stored in the database as the template. At the query stage, the visible and thermal face images undergo the identical process, and the resulting hash is matched against the template in the database (see Fig. 3(b));
3) The visible and thermal face features are hashed and enrolled in the gallery independently. At the query stage, the query visible and thermal face features are hashed to output the tuple (V', T'), which is then matched against the templates (V, T). Two matching scores are fused by taking the average, and the final decision (viz., yes/no) is made based on the average score and the predefined threshold value (see Fig. 3(c)).

V. Experiment

In this section, the feature extraction of thermal and visible images are first detailed. The accuracy of SMDSH is evaluated on thermal and visible face images. Next, the parameter optimization for NMDSH is performed. Finally, the performance of different application scenarios are reported.

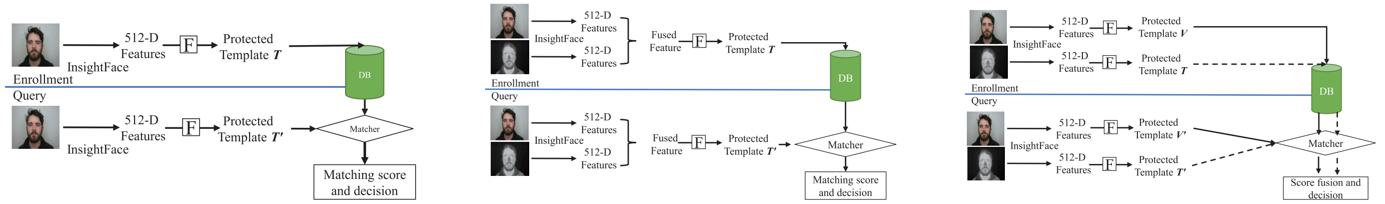
A. *Feature extraction*

For experiment purposes, the InsightFace face feature extraction method [9] is considered, and its open source codes from GitHub [17] are utilized. Visible and thermal images are co-registered by edge-based image registration approach according to [18]. The images are cropped to 112×112 so that they can be fed into the deep model for feature extraction. InsightFace model [19] pre-trained by MS-Celebrities [20] is employed to generate the 512-dimension embedding for both visible and thermal face images.

Here, we utilized the visible face images to train and build the deep model. The model is then deployed to extract features from thermal images. The results in the following experiment suggests that such deep model can also achieve satisfactory accuracy performance in face recognition for thermal images.

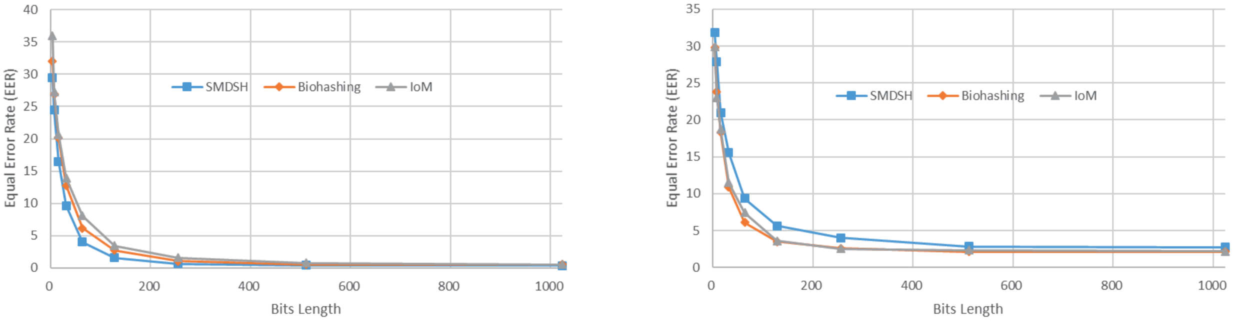
B. *SMDSH bits length optimization*

The effect of SMDSH bit length on accuracy for single modality system is investigated using equal error rate (EER). Specifically, EER is the rate at which both False Rejection Rate (FRR) and False Acceptance Rate (FAR) are equal, where lower EER indicates better system performance, and vice versa. Images of neutral face will be taken as gallery, and another emotion (e.g., emotion-happy, denoted as N-EH) will be taken as a probe. EER will be calculated for each case of variations. Take N-vs-EH as an example, firstly, 50 images of emotion-neutral are enrolled as the gallery subjects, then 50 images of emotion-happy are taken as probes. The N and EH features from the same person will be matched to generate the genuine score (i.e., 50 scores), while the EH and N features from different person will be matched to generate the imposter scores (i.e., $C_{50}^{2} = 1,225$ pairs). The average EER results are shown in Fig. 4, but due to space limitation, only EH, ESP, PR, OHE, LD and LR are reported. Results suggest that, for both visible and thermal faces, the accuracy performance stabilizes when the SMDSH bit length is ≥ 1024. For the rest of the discussion, unless specified otherwise, the length of 1024 bits is considered. To further investigate, the EER scores for matching emotion-neutral to other variations using SMDSH are recorded in Table II. Results suggest that the performance after applying SMDSH is slightly inferior (but still comparable) to that of the original features for the visible image scenario.



(a) Traditional visual image FR system (b) Feature fusion FR system (c) Score fusion FR system

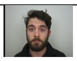
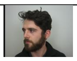
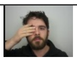
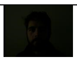
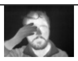
Fig. 3: Cancelable facial recognition systems. The boxed letter F in each figure refers to a transformation function (e.g., NMDSH), where the generated template is stored in the database denoted as DB. (a) shows a traditional FR system without fusion; (b) shows a FR system with feature fusion and hashing protection, and; (c) shows a FR system with hashing protection and score-level fusion.



(a) Visible face (b) Thermal face

Fig. 4: EER (in %) vs bit length on EURECOM VIS-TH dataset.

TABLE II: EER of matching neutral face to other variations with SMDSH for bit lengths of 1024.

Image example												
Image spectral	Visible						Thermal					
Gallery-Probe[a]	N-EH	N-ESP	N-PR	N-OHE	N-LR	N-LD	N-EH	N-ESP	N-PR	N-OHE	N-LR	N-LD
Original features	0.04	1.98	10.39	4	2.06	34.02	0.33	2.02	19.8	14.14	4.04	4.45
SMDSH	0.09	2.41	14.97	4.03	1.52	32.73	0.05	3.22	16.56	13.75	4.05	4.5

[a] N represents Neutral while EH, ESP and etc. correspond to other categories in Table I.

TABLE III: EER (in %) and inter-class variations under different nonlinear rate α values with NMDSH on visible face.

	SMDSH	NMDSH									
α	-	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
EER	1.3	1.12	1.25	1.16	1.17	1.63	3.9	5.52	11.54	14.29	18.13
$V[d-]$	903.23	893.7	894.01	880.58	691.24	455.67	327.86	298.51	299.17	297.37	301.53

C. Robustness of NMDSH against SA

As discussed in our previous work [10], SMDSH may be vulnerable to similarity-based attack although a many-to-one mapping function is employed. Hence, NMDSH is proposed to withstand SA. The inter-class variation [13], denoted by $V[d-]$, is utilized here. Note that an ideal system should achieve small inter-class variation while maintaining low EER value as indicated in [13]. Here, the value of $V[d-]$ in Hamming distance and EER under different nonlinear rate α are computed by using emotion related samples. The result is shown in Table III. It can be seen that large value of α can lead to a small inter-class variation, but a large α will lead to a drastic drop in accuracy. Therefore, it is necessary to strike a balance where the system preserves the accuracy while achieving small inter-class variation. It is also observed that $\alpha = 0.5$ is optimal for EURECOM visible face. When $\alpha = 0.5$ is set, EER does not degrade much and $V[d-]$ reaches an optimal value.

TABLE IV: EER (in %) of different scheme scenarios with SMDSH and NMDSH (value in parenthesis) for bit lengths of 1024.

<table>
<tr><th rowspan="2"></th><th colspan="2">Scenario</th><th rowspan="2">N-EH</th><th rowspan="2">N-ESP</th><th rowspan="2">N-PR</th><th rowspan="2">N-OHE</th><th rowspan="2">N-LR</th><th rowspan="2">N-LD</th></tr>
<tr><th>Gallery</th><th>Probes</th></tr>
<tr><td rowspan="2">Single Modality</td><td>visible</td><td>visible</td><td>0.05 (0.03)</td><td>2.69 (3.18)</td><td>12.58 (18.73)</td><td>4.02 (4.12)</td><td>1.62 (2.00)</td><td>31.83 (34.8)</td></tr>
<tr><td>thermal</td><td>thermal</td><td>0.11 (0.10)</td><td>3.01 (2.56)</td><td>17.52 (20.78)</td><td>13.56 (15.18)</td><td>3.81 (4.25)</td><td><b>5.08 (5.36)</b></td></tr>
<tr><td rowspan="3">Multi-Modalities[b]</td><td>visible</td><td>thermal</td><td>47.57 (51.7)</td><td>50.49 (50.7)</td><td>48.96 (51.5)</td><td>50.68 (52.2)</td><td>49.27 (51.4)</td><td>50.13 (48.14)</td></tr>
<tr><td>visible||thermal</td><td>visible||thermal</td><td>0.00 (0.00)</td><td>1.86 (0.97)</td><td>13.70 (<b>13.97</b>)</td><td>2.47 (3.47)</td><td>0.73 (<b>0.65</b>)</td><td>7.23 (7.68)</td></tr>
<tr><td>visible&thermal</td><td>visible&thermal</td><td><b>0.00 (0.00)</b></td><td><b>1.44 (1.16)</b></td><td><b>12.83</b> (15.72)</td><td><b>2.41 (3.27)</b></td><td><b>0.47</b> (0.96)</td><td>6.76 (7.19)</td></tr>
</table>

[b] || represents feature fusion by concatenation, and & represents score level fusion by averaging two matching scores.

D. Accuracy for different application scenarios

In a typical face biometric recognition system (verification system), several scenarios are possible and some representative scenarios are listed in Table IV. Here, || represents feature fusion by concatenation, and & represents score level fusion by averaging two matching scores. According to the discussions in Section II, we test the accuracy performance on two scenarios, namely, single modality and multi-modalities. Results in Table IV suggest that the highest accuracy is achieved by score level fusion of visible and thermal images. However, under the poor illumination situation (i.e. N-LD), single modality of thermal image provides the best results.

It is noteworthy that visible image features in the gallery do not match well with the thermal image features, which is evidenced by the results in Table IV (see first row under multi-modalities). This observation suggests that features extracted from thermal and visible images, although using the same deep model, show different characteristics. Finding a universal feature extractor for both visible and thermal images is still a challenging task. However, we can conclude that the deep model trained by visible images (e.g., InsightFace in this work) can also be utilized to extract features from the thermal face images, and the performances appear to be promising. For the deep model utilized in this work, it is still challenging to achieve satisfactory performance on pose variation images. In addition, for thermal image, it is also non-straightforward to generate discriminative features on occlusion variations, since LWIR can only capture the heat emitted from skin under no occlusion situation. Nevertheless, multi-modalities score level fusion on the transformed domain shows comparable performance in general. This suggests that fusion of visible and thermal can be a promising solution to address the issues discussed above.

VI. Conclusions

In this paper, we proposed a secure visual-thermal fused FR system using the newly developed VIS-TH face dataset by EURECOM. Moreover, a non-invertible transformation function is adopted to hash the deep features extracted by a pre-trained deep model. One cross-modality and two multi-modalities schemes with hashing function are explored. Results suggest that the utilization of thermal face can overcome the drawbacks of using visible face under limited or unfavorable illumination condition. Specifically, both score fusion and feature fusion show improvement in accuracy, but each type of fusion scheme can be applied under different situations. On one hand, feature fusion generates only one feature vector and a binary protected template as the final output, hence being economic on storage space. On the other hand, score level fusion achieves higher accuracy, and it is practical when some modalities are absence, e.g., no visible face available. It is noteworthy that although the InsightFace model is trained by using visible facial images, the model still exhibits comparable performance when applied on thermal images. We also demonstrate that NMDSH can preserve the performance of the FR system while offering extra layer of protection to face features.

As future work, we shall investigate into transfer learning for improving the accuracy of deep models on thermal face. In addition, we are also working on universal facial landmarks for both visible and thermal facial images, which potentially leads to cross-spectral matching.

References

[1] "Personal Identification and Descriptions," *Nature*, vol. 38, pp. 201–202, June 1888.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[3] 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*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[4] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1701–1708.

[5] Mei Wang and Weihong Deng, "Deep Face Recognition: A Survey," *arXiv:1804.06655 [cs]*, Apr. 2018, arXiv: 1804.06655.

[6] Guangcan Mai, Kai Cao, Pong C. Yuen, and Anil K. Jain, "On the Reconstruction of Face Images from Deep Face Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2018.

[7] Karthik Nandakumar and Anil K Jain, "Biometric template protection: Bridging the performance gap between theory and practice," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 88–100, 2015.

[8] Khawla Mallat and Jean-Luc Dugelay, "A benchmark database of visible and thermal paired face images across multiple variations," in *International Conference of the Biometrics Special Interest Group, BIOSIG 2018, Darmstadt, Germany, September*. LNI, pp. 199 – 206, GI / IEEE.

[9] Jiankang Deng, Jia Guo, and Stefanos Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," *arXiv:1801.07698 [cs]*, Jan. 2018, arXiv: 1801.07698.

[10] Dong XingBo, Wong KokSheik, Jin Zhe, and Dugelay Jean-luc, "A cancellable face template scheme based on nonlinear multi-dimension spectral hashing," in *7th IAPR International Workshop on Biometrics and Forensics - IWBF2019*. IEEE.

[11] Zhe Jin, Meng-Hui Lim, Andrew Beng Jin Teoh, and Bok-Min Goi, “A non-invertible Randomized Graph-based Hamming Embedding for generating cancelable fingerprint template,” *Pattern Recognition Letters*, vol. 42, pp. 137–147, June 2014.

[12] Yair Weiss, Rob Fergus, and Antonio Torralba, “Multidimensional spectral hashing,” in *European Conference on Computer Vision*. Springer, 2012, pp. 340–353.

[13] Yanzhi Chen, Yan Wo, Renjie Xie, Chudan Wu, and Guoqiang Han, “Deep Secure Quantization: On secure biometric hashing against similarity-based attacks,” *Signal Processing*, vol. 154, pp. 314–323, Jan. 2019.

[14] Mritunjay Rai, Tanmoy Maity, and R K Yadav, “Thermal imaging system and its real time applications: a survey,” *Journal of Engineering Technology*, vol. 6, no. 2, pp. 14, 2017.

[15] Thirimachos Bourlai and Bojan Cukic, “Multi-spectral face recognition: Identification of people in difficult environments,” in *2012 IEEE International Conference on Intelligence and Security Informatics*, Washington, DC, USA, June 2012, pp. 196–201, IEEE.

[16] Lin Hong, Anil K Jain, and Sharath Pankanti, “Can multibiometrics improve performance?,” in *Proceedings AutoID*. Citeseer, 1999, vol. 99, pp. 59–64.

[17] “Insightface source code,” https://bit.ly/2LTTjj3.

[18] Yong Sun Kim, Jae Hak Lee, and Jong Beom Ra, “Multi-sensor image registration based on intensity and edge orientation information,” *Pattern Recognition*, vol. 41, no. 11, pp. 3356–3365, Nov. 2008.

[19] “Insightface pre-trained model,” https://bit.ly/2RnF8WR.

[20] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao, “MS-Celeb-1m: A Dataset and Benchmark for Large-Scale Face Recognition,” *arXiv:1607.08221 [cs]*, July 2016, arXiv: 1607.08221.