2017 UKSim-AMSS 19th International Conference on Modelling & Simulation

Thermal Infrared Face Recognition: A review

Krishna Rao Kakkirala
TCS Innovation Labs
TATA Consultancy Services
Hyderabad, India
krishna(dot)kakkirala(at)tcs(dot)com

Srinivasa Rao Chalamala
TCS Innovation Labs
TATA Consultancy Services
Hyderabad, India
chalamala(dot)srao(at)tcs(dot)com

SantoshKumar Jami
TCS Innovation Labs
TATA Consultancy Services
Hyderabad, India
santoshkumar(dot)jami(at)tcs(dot)com

***Abstract*—Visible light face recognition has been well researched but still its performance is limited by varying illumination conditions. Illumination conditions are major source of the uncertainty in face recognition systems performance when it is used in outdoor setting. In order to augment the performance of visible light face recognition system's performance, infrared facial images as a new modality has been used in the literature. The other intriguing factor in using the infrared, especially thermal infrared imaging for facial recognition is the night time surveillance with little or no light to illuminate the faces. Thermal facial recognition is used in several covert military applications. The covert data acquisition has its own environmental constraints. The choice of infrared makes the system less dependent on external light sources and more robust with respect to incident angle and light variations.**

In this paper we focus on reviewing recent developments in thermal IR face recognition and suggest the approaches that could be useful in matching visible and thermal IR images.

Keywords–Thermal IR; face recognition; cross modality matching; deep neural networks

I. INTRODUCTION

One of the major limitations in visible light face recognition is the illumination dependency problem which significantly decreases the efficiency of the system, especially, in the outdoor and night vision applications.

To overcome these limitations several solutions have been investigated. One of the solutions for the mentioned limitation is using infrared facial images. Recently night vision devices such as passive infrared cameras are introduced. These passive thermal infrared cameras capture the radiation emitted by objects between wavelengths of 3-14 μm. Using this thermal infrared imagery for face recognition has been growing in critical security applications. Several methods have been proposed for thermal infrared face recognition. Some of the methods used in visible light face recognition are used on thermal IR face recognition. These include local feature based methods which extract distinctive local structural information. In fusion based methods images or features or decision fused when both the visible and thermal images are available. Cross modality based methods match between visible and thermal imagery using distinctive features. Deep neural network (DNN) based methods find mapping between thermal and visible imagery and features.

The infrared spectrum is divided into four bandwidths: Near IR (NIR), Short wave-IR (SWIR), Medium wave IR (MWIR) and Long wave IR (Thermal IR). Thermal IR has received the most attention due to its robustness among others. Thermal IR sensors measure the emitted heat energy from the object and they dont measure the reflected energy.Thermal spectrum has some advantages. Thermal images of the face scan be obtained under every light condition even under completely dark environments. Thermal energy emitted from the face is less affected to the scattering and the absorption by smoke or dust. Thermal IR images also reveal anatomical information of face that makes it capable to detect disguises [1].

This paper aims to provide survey on recent progress in thermal IR face recognition research. It contains few new references not found in previous surveys.

This paper prescribes new approaches that could be explored for cross modality face matching between visible and thermal IR images.

II. II. LITERATURE REVIEW ON THERMAL FACE RECOGNITION

1) Visible light face recognition: The objective of any face recognition system is to find and learn features that are distinctive among people. While it it important to learn the distinctive feature, it is also important to minimize the differences between images of a same person captured under several conditions. These variations included, distance between camera and person, lighting changes, pose changes and occlusions.

Visible light face recognition has been well researched and the performance reached acceptable levels but still the variation in pose and illuminations are a problem which is limiting the algorithms reach 100% accuracy. The visible light face recognition systems are based on local and global features of the face that are discriminative under controlled environment. Pose robust visible light algorithms use common subspace learning in which the images or features are transformed into a subspace where the intra-class variations are minimized and inter class variations are maximized. The subspace or the manifold must allow better classification of the facial images.

978-1-5386-2735-8/17 $31.00 © 2017 IEEE
DOI 10.1109/UKSim.2017.38

Recently deep learning have made rapid progress in face recognition. The important ingredient being the convolutional filters. Various deep architectures have been proposed to improve the face recognition performance under various factors such as pose, low resolution distance, illuminations. The disadvantage of these methods being they require huge amount of data for training. This has been partially overcome by using transfer learning. But still they require good amount of data and learning time. Some people used deep neural networks (DNN) only for computing the features that can be readily classified using various classifiers for better performance. This reduced the training time as the DNN required to just compute the features.

These visible light face recognition methods completely fail when there is no sufficient light or complete darkness. Infrared imaging helps in capturing the facial images even under complete darkness.

A. *Thermal IR Face recognition: Approaches*

Use of Thermal IR imaging in face recognition enable it to be invariant to even extreme illumination changes. Face recognition using the visible spectrum may not work properly in low lighting conditions.

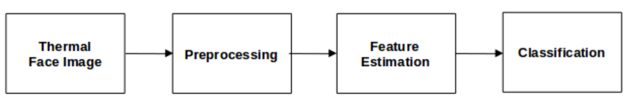

Figure 1. Infrared Face Recognition

Thermal IR sensor measures the heat energy radiation, not the reflectance from the objects hence thermal IR imagery is less sensitive to the variations in facial appearance caused by illumination changes. Figure 1 shows a generic thermal IR face recognition system which is like any visible light face recognition system computes distinctive features and learns them for recognition. Figure 2 shows the visible light and thermal images are merged together before feature extraction. This merged image represents both the modalities and can recognize faces from any modality during testing. Figure 3 shows how a visible light and thermal features are concatenated for improvement in recognition with images of any modality is available.

Thermal infrared face recognition uses most of the visible light face recognition work. Since both the test and training sets are Thermal IR these systems perform similar to visible light face recognition systems. Since Thermal IR images can be captured at various spectra there is the performance difference for each spectra could be different.

The Thermal IR imagery can be captured at MWIR or SWIR. The performance of a method could be different for each of these spectra as spectra can suppress useful information also add noise due to their very nature.

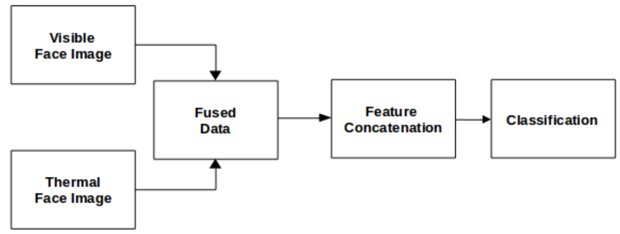

Figure 2. Cross modality face recognition using Visible and Thermal IR images

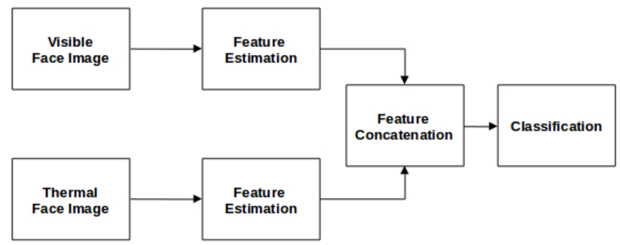

Figure 3. Heterogeneous face recognition using concatenated features

In situations where both the visible and thermal imagery is available for training then the face recognition system can use the both the information for better learning. As seen in figure 3 the feature of both the visible and thermal IR spectra can be concatenated to model the faces. The images of visible and infrared can be merged or fused to come up with common image that can include information of both the modalities of the imagery so that given a test image whether it is visible or infrared the system works reliably. The recognition be inferior when the test set has only single modality.

Figure 4 shows a thermal IR face recognition system, that instead of merging the features or the imagery it projects them into a common subspace where the features of the a same person are projected such that they are close by and are well separated from the other. Canonical correlation analysis (CCA) and similar methods help in projecting the features into this common subspace.

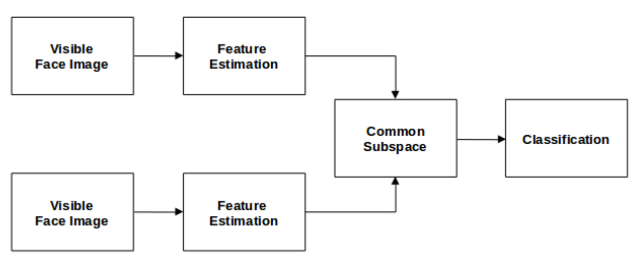

Figure 4. Heterogeneous face recognition common subspace learning

Cross modal face matching between the thermal and visible spectrum is a much desired capability for the applications like night-time surveillance. A thermal FR system must capture a highly non-linear relationship between the two modalities by using a deep neural network.

Recently deep learning made waves in face recognition by showing performance close to humans. Some researchers experimented with finding common features space for thermal IR face recognition. Figure 5 shows a system where the DNN learns modality mapping between visible light facial features and thermal IR features. Once the DNN learns this mapping given a test thermal IR image, the DNN would be able to map it to a corresponding visible light images. There are other ways how the deep learning can be used for Thermal IR face recognition.

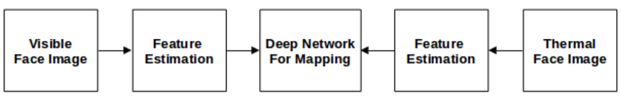

Figure 5. Thermal IR and Visible features mapping using DNN

As in visible light face recognition system the thermal IR face recognition systems require several steps that are necessary for effective feature extraction. The steps include face detection, localization, pre-processing and feature extraction.

B. Thermal IR face recognition: Pre-processing

Pre-processing of face images is necessary step for extraction of effective features. After face detection the face images are cropped and normalized. Low pass filtering helps in removing unwanted noise in the images and to remove illumination variations. The normalized and cropped visible and thermal face images are then subjected to Difference of Gaussian (DOG) filtering, which is a common technique to remove illumination variations for visible face recognition [2]. It has been used in visible face recognition to a large extent. A DOG filter image is constructed by convolutions of an original image and DOG filter. For images a 2-dimensional DOG filter is used. DOG filtering, not only reduces illumination variations in the visible facial imagery, but also reduces local variations in the thermal imagery that arises from varying heat/temperature distribution of the face. Therefore, DOG filtering helps to narrow the modality gap by reducing the local variations in each modality while enhancing edge information. The σ of the Gaussian filter is selected carefully to obtain components that benefit face recognition and suppress those frequencies that negatively affect face recognition. Methods using local features extraction benefit from DOG filtering extensively as it will remove unwanted noise and enhances the required features such as edges. Same DOG filtering can be used for thermal IR images with appropriate σ values.

C. Thermal face recognition: Local Features

Initially thermal IR face recognition research used the methods developed for visible light face recognition. Diego A. Socolinsky et al. of Equinox corporation studied appearance-based face recognition algorithms on visible and LWIR imagery [3]. They considered PCA (principle component analysis), LDA (linear discriminant analysis) and ICA analysis tools for evaluating the recognition performance. Visible images are illumination compensated and Thermal images were radiometric calibrated. A brief description of datasets has been given in section V

They used subspaces of 100- dimensional for PCA, LFA and ICA, and the LDA. The authors found recognition performance on visible imagery, regardless of algorithm, that was worst for pairs where both illumination and facial expression vary between the training and testing sets. The authors noted from these results that the recognition performance is better with LWIR over visible imagery [4].

The use of the LBP operator in face recognition was introduced in [5] and different extensions from the original operator have appeared afterwards [6]. Local Binary Patterns (LBP) used as a means of summarizes local gray level structure of an image[36]. LBPs have proven to be highly discriminative features for texture classification [7] and they are resistant to lighting effects in the sense that they are invariant to gray level transformations. Many facial regions are relatively uniform and it is legitimate to investigate whether the robustness of the features can be improved in these regions.

LWIR face images are collected by equinox [26]. This dataset comprises of 40 frames of 91 persons each uttering vowels with frontal pose, with various lighting settings. Also some images are taken with the subjects are putting more expressions. The study has been conducted with and without glasses. The authors compared the LBP performance against LDA and found that average performance of both LBP and LDA are similar when there are no glasses. The performance degrades with eye glasses. The recognition is performed using a nearest neighbor classifier in the computed feature space, using similarity measures like histogram intersection, log-likelihood statistic and Chi square etc., [2].

The authors in [10], [11], [12] compared various thermal face recognition methods over Equinox and UCH thermal face datasets and found that Weber Local Descriptor(WLD) show better performance on Equinox dataset even with eye glasses. Appearance based methods are not an option for faces with facial expressions or artifacts. WLD performs well even with simple gallery image. SIFT and SURF gives us the best performance on UCHThermal face dataset which is the most robust against rotations and facial expressions.

In recent years approaches like wide baseline matching have become increasingly popular and have experienced impressive development. Typically, these methods are based on local interest points which are extracted independently from both a test and a reference image, and then characterized by invariant descriptors, and finally the descriptors are matched until a given transformation between the two images is obtained. Lowes system [13], using SIFT descriptors and a probabilistic hypothesis rejection stage, is a popular choice for implementing object recognition systems, given its recognition capabilities, and near real-time operation. In addition the SIFT features can be used to register visible and thermal face images.

D. Thermal face recognition: Fusion

The fusion of thermal and visible images can be done in image level, feature level (see figures 2 and 3), match score level and decision level. In those the simplest form of fusing visible and infrared images are by concatenating the feature vectors. Some authors proposed fusing the images in Eigen space. The authors of [14] fused both at image level and decision level. The authors [15] fused images by transforming them into wavelet domain and trained on 2V-GSVM. [16]

Bebis et. al. [8] studied the sensitivity of thermal IR imagery to facial occlusions caused by eyeglasses. Specifically, their experimental results illustrate that recognition performance in the IR spectrum degrades seriously when eyeglasses are present in the probe image but not in the gallery image and vice versa. To address this serious limitation of IR, the authors fused IR with visible imagery. In contrast, fusion at multiple resolution levels allows features with different spatial extend to be fused at the resolution at which they are most salient. In multi-resolution fusion [24] both visible light and thermal images are multiscale transformed then a composite multiscale representation is constructed from these. During this construction some specific fusion rules are used. The fused image is obtained by taking an inverse multiscale transform. Multiscale face representations have been used in several systems [24].

E. Thermal face recognition: Deep Neural Networks

An alternative method for face recognition would be deep learning. Deep learning exhibited more advantages and has been applied to many different fields such as computer vision and pattern recognition. Please see Figure 5. Deep learning achieves the complex function approximation through a nonlinear network structure and shows the powerful learning ability. Compared with traditional recognition algorithms, deep learning combines feature selection or extraction and classifier determination into one step and can study features to reduce the workload for the manual design feature.

In [17] authors used a convolutional neural network (CNN) architecture for thermal face recognition. CNN is a kind of deep learning algorithm. In this thermal images are firstly obtained from the RGB-D-T face database applied to CNN network, which can automatically learn effective features from the thermal face data. Compared with other state-of-art methods such as LBP, HOG and moments invariant, this approach giving better recognition rates. They tested this method on RGB-D-T face database in which the images are collected from RGB camera, Kinect camera, and thermal imaging camera synchronously. Their method achieved the recognition rates 98%, 99.4%,and 100% on head rotations, expression variations and illumination variations respectively.

Due to a very large modality gap, thermal-to-visible face recognition is one of the most challenging face matching problem. Deep Perceptual Mapping (DPM) [18] is approach that captures the non-linear relation between these modalities by using Deep Neural Networks (DNN).

This network learns the non-linear mapping from thermal to visible spectrum while preserving the information that defines the person identity. After obtaining the mapping from visible to thermal domain, the mapped descriptors of the visible gallery images are concatenated together to form a long feature vector. The resulting feature vector values are normalized and then matched with the similarly constructed vector from the probe thermal image. University of Notre Dames UND collection was used in testing which contains 4584 images of 82 subjects distributed evenly in visible and thermal domain. This approach is given rank-1 Identification accuracy of 83.73% using all thermal images as probes and visible images in the gallery.

III. Limitation of Thermal IR Face Recognition

Thermal face recognition suffers from several problems. One of the major problem of the thermal spectrum is the occlusion due to opaqueness of eyeglasses to the spectra. Eye glasses cause occlusion leading to large portion of the face occluded, causing the loss of important discriminative information. Persons with identical facial parameters might have radically different thermal signatures. MWIR and LWIR images are sensitive to the environmental temperature, as well as the emotional, physical and health condition of the persons. Alcohol consumption changes the thermal signature of the persons which could lead to performance degradation.

IV. Datasets

A. UND Collection X1 [34]

- Collected by the University of Notre Dame
- Using a Merlin uncooled LWIR camera and a high resolution visible color camera
- Consisting of visible and LWIR images
- Acquired across multiple sessions for 82 distinct subjects
- The total number of images is 2,293

B. WSRI Dataset [35]

- The WSRI dataset was acquired by the Wright State Research Institute at Wright State University.

- Number of subjects 119 subjects
- 3-visible, 3- corresponding MWIR images
- Manually annotated fiducial points (left eye, right eye, tip of nose, and center of mouth)

C. Equinox Dataset [26]

- LWIR images collected by Equinox Corporation.
- This database is composed of 3 sequences of 40 frames from 91 persons,
- Three different lights source: frontal, left lateral and right lateral
- Captured while people were pronouncing the vowels standing in frontal position
- Three more images from each person with smile, frown and surprise expressions
- Repeated for those persons who wore glasses

V. Way forward for Thermal IR Face recognition

Subspace learning methods are one type of the most popular methods for cross-modality matching. They aim to learn a common subspace shared by different modalities of data, in which the similarity between different modalities of data can be measured. Unsupervised subspace learning methods use pairwise information to learn a common latent subspace across multi-modal data. They enforce pair-wise closeness between different modalities of data in the common subspace.

Restricted Boltzmann Machines (RBM) is an undirected graphical model that can learn the distribution of training data [23] to learn a joint latent space for matching images.

Besides CCA, Partial Least Squares (PLS) and Bilinear Model (BLM) are also used for cross-modal retrieval. These cross-modal methods map images with different modalities to a common linear subspace in which they are highly correlated.

Recently, novel deep neural network architectures attracted lot of attention of researchers to solve several difficult learning problems due to its improved state of the art performance in their respective domains like image classification [19], face recognition [20], image generation [21] etc., Various deep architectures can be used for cross modal face recognition to reduce the modality gap between two modalities, hence cross modal face recognition can be performed. One of them is usage of Siamese neural network [22] can be used to extract common features from two modalities, and any classifier can be used to classify the given probe face. Another deep architecture CorrNet [23] can be used to reduce modality gap between thermal and visible images for thermal to visible face recognition applications, where this network ensures that common representations of the two modality faces are correlated. See figure 5.

Inspired by the success of deep neural networks , a variety of deep multi-view feature learning methods have been proposed to capture the high-level correlation between multi-view data. Recently, a deep canonically correlated autoencoder (DCCAE) [32] is proposed by combining the advantages of the deep CCA and autoencoder based approaches. DCCAE consists of multiple autoencoders and optimizes the combination of canonical correlation between the learned representations and the reconstruction errors of the autoencoders.

Deep Canonical Correlation Analysis (DCCA), [31] a deep learning method can learn complex nonlinear projections for different modalities of data such that the resulting representations are highly correlated. A metric learning approach for cross modal matching, which considers both positive and negative constraint. Restricted Boltzmann Machines (RBM) is an undirected graphical model that can learn the distribution of training data.

VI. Conclusion

In this paper we reviewed some important papers that exists in the literature for thermal IR face recognition. Both the local and global features are used for the recognition. Local features based methods performance better than global features. When the both the visible and thermal imagery is available fusion based methods and DNN are used to find common feature space. MWIR face recognition (VTD Dataset) performs comparatively with visible light face recognition[25]. MWIR performs better than visible face recognition when images acquired in complete darkness.

Texture based methods preform well with both MWIR, visible for face recognition. Appearance based methods performs better in MWIR spectrum because MWIR images are illumination invariant. LWIR imagery is robust to illumination variations and have less intra class variations.

Cross spectral matching could help when the data is not sufficient and the training dataset doesn't contain thermal IR images. Most of the existing systems has only visible images captured for training. In order to use existing systems and enable recognition in dark and outdoor scenarios it is necessary to develop systems that can match between thermal and visible images. Cross-spectral matching is challenging than intra-spectral matching though the rank-1 accuracies of cross-spectral matching is low they can be used for face verifications rank-N accuracies are higher. We suggested few cross modality matching approaches in section V.

When imagery of both the modalities is available it is imperative to force the systems to learn both the modalities for better recognition rates. Common Representation Learning (CRL), wherein different modalities of the data are embedded in a common subspace, is receiving a lot of attention recently. Canonical Correlation Analysis is one such approach. Deep neural networks that can learn correlated common representations could be of interest to the researchers in the coming future.

REFERENCES

[1] R. S. Ghiass, O. Arandjelovic, A. Bendada, and X. Maldague, ”Infrared face recognition: A literature review.” IJCNN, 2013.

[2] Xiaoyang Tan, B Triggs, ”Enhanced Local Texture Feature Sets for Face Recogniton Under Difficult Lighting Conditions”, IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1635-1650, 2010.

[3] D. A. Socolinsky, A. Selinger, ”A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery”, IEEE Proc. 16th Int. Conf. on Pattern Recognition, vol. 4, pp. 217-222, 2002.

[4] H. Mendez, C. SanMartn, J. Kittler, Y. Plasencia, E. Garca, Face recognition with LWIR imagery using local binary patterns , LNCS5558(2009), 327336.

[5] Ahonen, T., Hadid, A., Pietikinen, M.: Face recognition with local binary patterns In: Pajdla, T., Matas, J. (eds.) Computer Vision, ECCV 2004 Proceedings, LNCS 3021. Springer, Heidelberg (2004)469-481.

[6] Marcel, S., Rodriguez, Y., Heusch, G.: On the Recent Use of Local Binary Patterns for Face Authentication. In: International Journal on Image and Video Processing Special Issue on Facial Image Processing (2007).

[7] SS Ghatge, VV Dixit, Robust Face Recognition under Difficult Lighting Conditions ,International Journal of Technological Exploration and Learning (IJTEL) Volume 1 Issue 1 (August 2012)

[8] Bebis G, Gyaourova A, Singh S, Pavlidis I. Face recognition by fusing thermal infrared and visible imagery. Image and Vision Computing. 2006 Jul 1;24(7):727-42.

[9] Ketki D. Kalamkar, Prakash.S. Mohod, Feature extraction of Heterogeneous face images using SIFT and MLBP algorithm, International Conference on Communications and Signal Processing (ICCSP), 2015

[10] G. Hermosilla, J. Ruiz-del-Solar, R. Verschae, and M. Correa, A comparative study of thermal face recognition methods in unconstrained environments,” Pattern Recognit., vol. 45, no. 7, pp. 2445-2459, 2012.

[11] B. Klare and A. Jain. Heterogeneous face recognition using kernel prototype similarities. IEEE Trans. Pattern Analysis and Machine Intelligence, 35(6):1410-1422, June 2013

[12] T. Bourlai, A. Ross, C. Chen, and L. Hornak. A study on using midwave infrared images for face recognition. In Proc. SPIE, 2012.

[13] D. Lowe, Distinctive image features from scale invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91110.

[14] R. Singh, M. Vatsa, and A. Noore, ”Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition,” Pattern Recognition , Vol.41, pp.880-893, 2008.

[15] R. Singh, M. Vatsa, and A. Noore, ”Hierarchical fusion of multispectral face images for improved recognition performance,” Information Fusion , Vol.9, pp.200-210, 2008.

[16] Y Bi, M Lv, Y Wei, N Guan, W Yi, Multi-feature fusion for thermal face recognition, Infrared Physics & Technology, 2016 Elsevier

[17] Zhan Wu ; Min Peng ; Tong Chen, Thermal Face Recognition Using Convolutional Neural Network International Conference on OptoElectronics and Image Processing (ICOIP), 2016

[18] Sarfraz, M., Stiefelhagen, R. Deep perceptual mapping for thermal to visible face recognitionIn: BMVC (2015)

[19] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009.

[20] Karen Simonyan, Andrew Zisserman,”Very deep convolutional networks for large-scale image recognition”, IEEE Computer Vision and Pattern Recognition (CVPR), 2015

[21] Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu,” Pixel Recurrent Neural Networks”, IEEE Computer Vision and Pattern Recognition (CVPR), 2016

[22] Bromley, J. W. Bentz, L. bottou, I. Guyon, Y. LeCun, C. Moore, E. Sckinger, R. Shall, ”Signature verification using a siamese time delay neural network”, Int. J. Pattern Recognit. Artificial Intell., vol. 7, no. 4, pp. 669-687, Aug. 1993.

[23] S. Chandar, M. M. Khapra, H. Larochelle and B. Ravindran, ”Correlational Neural Networks,” in Neural Computation, vol. 28, no. 2, pp. 257-285, Feb. 2016.

[24] Manjunath, B.S., Chellappa, R. and von der Malsburg, C., 1992, June. A feature based approach to face recognition. In Computer Vision and Pattern Recognition, 1992. Proceedings CVPR'92., 1992 IEEE Computer Society Conference on (pp. 373-378). IEEE.

[25] Bourlai, T., Ross, A., Chen, C. and Hornak, L., 2012, May. A study on using mid-wave infrared images for face recognition. In SPIE Defense, Security, and Sensing (pp. 83711K-83711K). International Society for Optics and Photonics.

[26] http://www.equinoxsensors.com/products/HID.html

[27] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. R. Lanckriet, R. Levy, and N. Vasconcelos, A new approach to cross-modal multimedia retrieval, in International conference on Multimedia. ACM, 2010, pp. 251260

[28] R. Rosipal and N. Kramer, ”Overview and recent advances in partial least squares,” in subspace, latent structure and feature selection, Springer, 2006, pp. 3451.

[29] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, ”Generalized multiview analysis: A discriminative latent space,” in Computer Vision and Pattern Recognition. IEEE, 2012, pp. 21602167

[30] A. Lu, W. Wang, M. Bansal, K. Gimpel, and K. Livescu, Deep multilingual correlation for improved word embeddings, in HLT-NAACL , 2015, pp. 250256.

[31] F. Yan and K. Mikolajczyk, Deep correlation for matching images and text, in CVPR , 2015, pp. 34413450.

[32] W. Wang, R. Arora, K. Livescu, and J. A. Bilmes, On deep multi-view representation learning, in ICML , 2015, pp. 1083 1092.

[33] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, Multimodal deep learning, in ICML , 2011, pp. 689696.

[34] https://sites.google.com/a/nd.edu/public-cvrl/data-sets

[35] https://wsri.wright.edu/

[36] Chalamala, Srinivasa Rao, Santosh Kumar Jami, and B. Yegnanarayana. ”Enhanced face recognition using Cross Local Radon Binary Patterns.” Consumer Electronics (ICCE), 2015 IEEE International Conference on. IEEE, 2015.