A Survey on Heterogeneous Face Matching: NIR Images to VIS Images

Sandhya R.Wadhavane[1] Prof.Dr.S.M.Kamlapur[2]
[1]M.E. Student [2]Professor
[1,2]Department of Computer Engineering
[1,2]K.K.W.I.E.E.R., Nashik Savitribai Phule Pune University, Maharashtra, India

Abstract— Human face matching plays an important role in applications such as criminal identification, banking, video surveillance and credit card verification. Face images are captured by different acquisition systems like visible light cameras and infrared imaging devices. Visible light cameras capture visible light images while near infrared images are captured by infrared imaging devices. These images are called heterogeneous images as they have different image formation characteristics like qualities of cameras, sensors used in cameras and different environment conditions. Matching heterogeneous images is subject of the research. This survey provides a review of different existing techniques for heterogeneous face matching. These techniques commonly categories into two parts namely Invariant feature based techniques and Common space based techniques. Moreover, the datasets available and performance parameters for evaluation are also discussed.

Key words: Acquisition System, Heterogeneous Face, Near Infrared (NIR) Face Image, Visible (VIS) Face Image

I. INTRODUCTION

Face images are captured by different acquisition systems like visible light cameras capture visible light (VIS) images while near infrared (NIR) images are captured by infrared imaging devices. The images can be captured at daytime or nighttime so illumination conditions also differ. Such types of images are called as heterogeneous images and heterogeneous face matching refers to matching face images across different modality. The heterogeneous (NIR-VIS) face images are shown in figure 1. In many face matching system, one of the challenging task is to match the heterogeneous face images. Many applications such as E-passport, video surveillance, and photo based identification requires heterogeneous face matching because, in these applications probe face images and gallery face images are of different modalities such as NIR (Near Infrared),VIS (Visible Light), TIR (Thermal Infrared),3D images, Sketch images. There can be following different scenarios of heterogeneous face matching:

- NIR face image vs. VIS face image
- NIR face image vs. TIR face image
- VIS face image vs. Sketch
- VIS face image vs. 3D face image

The key problem in heterogeneous (NIR-VIS) face matching is that images of the same object may differ in appearance due to different image modality. Typical heterogeneous face matching activity involves following three main steps:

1) Face image representation: Facial representation is nothing but the description of facial information. Face images are represented using feature based representation methods. Face representation involves face image normalization and image filtering. Normalization involves some basic activity like rotation, scaling, and cropping. Filters are used to remove the noise. There are different filters used in image processing such as Laplacian of Gaussian (LoG) Center-Surround Division Normalization (CSDN), Difference of Gaussian (DoG), Gaussian smoothing filter.
2) Feature extraction: After normalization, the next step is to extract features using some of the descriptors. There are different descriptors used in image processing for feature extraction like Scale Invariant Feature Transform (SIFT), Local Binary Pattern (LBP), Multi-scale Local Binary Pattern (MLBP), HOG descriptor, Micro Block-Local Binary Patten (MB-LBP), Coupled-Information Tree Encoding (CITE) and Gabor.
3) Matching: After extracting the features, similarity between the feature vectors are measured and compared using different methods.

Visible and near infrared face image matching is also called cross-spectral face matching, which is important for security application. Visible and near infrared face image matching offer a potential to face recognition in various illumination conditions, especially in nighttime security system. So research focuses on VIS-NIR face matching.

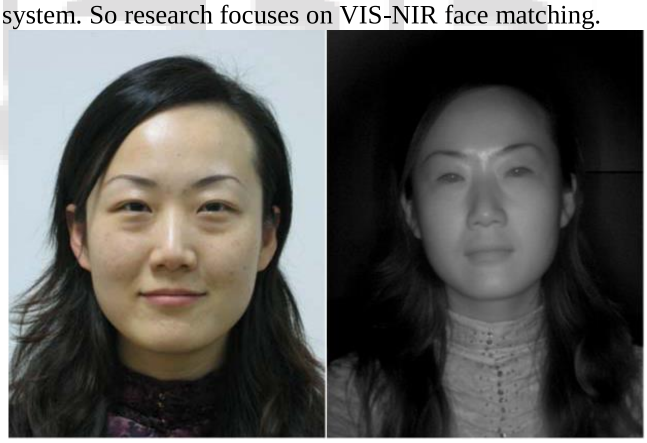

Fig. 1: VIS -NIR face image

II. RELATED WORK

One of the most difficult task in NIR-VIS face image matching is the appearance differences which are caused because of illumination variation. Lots of work has been done to tackle this problem. The existing heterogeneous face matching method for visible light to near infrared face image grouped into three categories:

A. *Invariant Feature Based*

These methods are useful to find the invariant features which are robust in illumination variation conditions. Generally it includes local binary patter (LBP) and Scale invariants feature transform (SIFT).

In many face recognition system gallery and probe faces are of different modality. So there is great difference

All rights reserved by www.ijsrd.com

between appearance of query and gallery face image therefore it becomes difficult to match both faces. To reduce the difference between VIS and NIR facial images Goswami et al. [10] introduced an effective preprocessing chain based on Gamma correction, Difference-of-Gaussian (DoG) filtering and contrast equalization. Then Local Binary Pattern (LBP) is used to extract invariant feature combine with linear discriminant analysis (LDA).

To address heterogeneous face matching problem Huang et al. [11] proposed another method called, modality invariant features (MIF). In this, three different modality-invariant features are extracted namely, sparse coefficients (SC), least square coefficients (LSC), and quantized distance vector (QDV). Then, represented these features with a sparse coding framework and sparse coding coefficients are used as the encoding for matching.

To address the illumination variation problems in NIR-VIS face matching Zhu et al. [16] designed a descriptor, called logarithm gradient histogram (LGH). It considers illumination direction, magnitude and spectral wavelength. This descriptor performs better than local binary pattern [5] and scale invariant feature transform descriptors. Training set is not required as it is feature-based approach

To maximize the correlation of the encoded face images between NIR and VIS face image and to reduce within-class variations Gong and Zheng [19] proposed a new feature descriptor. As it reduces within class variation, offers better performance than Histogram of Gradient (HOG), Local Binary Pattern (LBP) and Multiscale Local Binary Pattern (MLBP); but unlike the others it requires training.

In most of existing security systems people are registered only using VIS images, i.e. the registered VIS face do not have the corresponding NIR face in the system. That is corresponding pair of VIS-NIR images are not always available, so existing NIR-VIS face matching methods are not applicable. To address this problem Zhu et al [20] proposed a method called transductive heterogeneous face matching (THFM) which adapts the VIS-NIR matching learned from training with available image pairs.

To extract the common features from infrared face images and optical face images, a new learning-based face descriptor is proposed by Lei et al. [21], called Common Feature Discriminant Analysis (CFDA). This descriptor has greater accuracy than previous descriptors as it maximizes the correlations between infrared face images and optical face images. To encode optical and infrared face images designed a hyperplane based encoding method. To avoid the over-fitting, local feature based discriminant analysis (LFDA) based two level face matching method is applied for fast matching.

B. Projection based

These methods project the images from both modalities to a common subspace to reduce the appearance difference between them. It includes different methods such as linear discriminant analysis (LDA), canonical correlation analysis (CCA).

A new algorithm, Common Discriminant Feature Extraction (CDFE) for NIR-VIS face matching was proposed by Dahua Lin and Xiaoou Tang [1], where two linear mappings are learned to transform the samples from both NIR and VIS modalities to a common feature space. The algorithm is further extended into two non-linear approaches to deal with more complicated situations. The first extension is Kernelized Discriminant Feature Extraction (KCDFE), which is useful to extract the non-linear features and the second is multi-mode framework.

To learn more distinctive features of near infrared (NIR) image and visible light (VIS) image, Canonical Correlation Analysis (CCA) is used by Yi et al. [2]. In this, linear discriminant analysis (LDA) is used to extract features and to reduce the dimension of the feature vectors. For better classification, CCA learning is performed between features in LDA subspace instead of performing between images. This approach does not consider class label information in canonical correlation analysis process so it may drop some important information which is helpful for classification.

The main drawback of both the algorithms proposed by Tang and Yi is over-fitting. To overcome this drawback Liao et al. [5] proposed a new method, called Local Structure of Normalized Appearance (LSNA). In this approach, Difference-of-Guassian (DoG) filter is used to normalized heterogeneous face images .Then represent the features by applying Multi-scale Block LBP (MB-LBP) and for MB-LBP features selection Adaboost is used. Further, Regularized Linear Discriminant Analysis (R-LDA) is applied on training data to classify heterogeneous face images. Finally, heterogeneous face matching is performed using cosine distance measure

Klare and Jain [8] proposed a new approach for matching near infrared (NIR) face image to visible light (VIS) face image. This approach is based on method of Liao et al. [5], but it offers some improvements. They add Histogram of Oriented Gradient (HOG) feature descriptor to Local Binary Pattern (LBP) descriptor to extract more important information. To improve the generalization capability and to handle high dimensional feature vector, learn an ensemble of random LDA subspaces. Finally, matching is performed using Nearest Neighbor (NN) and Sparse Representation (SR) based matching.

Lei et al. presented a method for NIR and VIS face matching, called Coupled Spectral Regression (CSR) [4]. Similar to CDFE method, they used two projections to project the heterogeneous (NIR-VIS) data into a common discriminant feature subspace. The CDFE method [4] is time-consuming and performs well only on small data size. In order to improve the efficiency and reduce the time complexity, graph embedding and spectral regression based techniques combined with regularization techniques are used. Lei et al. later improve the CSR framework to construct the projections from different modality and to better exploit the locality information, called an Improved Coupled Spectral Regression (ICSR) [12]

C. Synthesis based

In this, synthesize one modality face image based on other modality and the synthesized image can then be used directly for homogeneous face matching. Typical methods include eigentransforms and local linear embedding (LLE).

A new method for heterogeneous (NIR-VIS) face mapping, proposed by Wang et al. [3], called face analogy

All rights reserved by *www.ijsrd.com*

based on analysis-by-synthesis framework. In this framework, NIR face image is transform in to VIS face image and then matching between synthesized image and original VIS image is performed. There are two steps in face analogy method namely local normalization and synthesis step. Local normalization operator constructs common local invariants of NIR and VIS face images. Although face analogy is used for NIR-VIS face mapping, it could be applicable for mapping VIS -3D face images.

Further improved synthesis of VIS image using locally-linear embedding (LLE) by Chen et al. [6], which is based on cross-domain dictionary learning. VIS images can be synthesized patch-by-patch by finding the best matching patch for each patch of the input NIR image. To more reliably match patches, illumination invariant LBP features are used to represent them. Finally matching between synthesized VIS image and original VIS image is performed using nearest neighbor classifier on the LBP representations of the synthesized images.

Xiong et al [15] proposed a new synthesis based method for NIR-VIS face mapping. To eliminate the influences of facial structure variations, a 3D model is used to perform pose rectification and pixel-wise alignment. Difference of Gaussian (DOG) filter is further used to normalize image intensities.

III. NIR-VIS FACE DATASET

Following are the NIR-VIS face image datasets.

A. CASIA HFB Dataset:

This includes total 100 subjects, in which 57 males and 43 are females each having 4 VIS and 4 NIR face images. Also, there are 3D images for each subject. In total, there are 800 NIR-VIS images and 200 3D images [7]

B. CASIA NIR-VIS 2.0:

This dataset includes 725 subjects each having 22 VIS and 28 NIR images. In total there are 36,250 images [18]

C. Cross Spectral Dataset:

It consists of 430 subjects each have at least one set of 3 poses (-10 degree / 0 degree / 10 degree). In total, there are 2,103 NIR images and 2,086 VIS images. [10]

D. PolyU NIR Face Dataset:

This dataset consist of 33,500 images with frontal face images, faces with expression, and pose variations of 335 subjects.[9]

E. LDHF Face Dataset:

It consist of both VIS and NIR face images of 100 different subjects, in which 70 are males and 30 are females. Images are captured at various standoff (60m, 100m, and 150m).[13]

IV. PERFORMANCE METRICS

The performance of the face matching system can be calculated using verification time and accuracy.

A. Verification Time:

Verification time is the duration of time that a system takes to make decision.

B. Matching Accuracy:

Matching accuracy is defined as follows:

$$\text{Matching Accuracy} = \frac{\text{Number of correct matches}}{\text{Total Number of testing images}} \times 100$$

V. CONCLUSION

Because of the decreasing cost of NIR acquisition devices, are becoming an important component in surveillance cameras.

Most heterogeneous (NIR-VIS) face matching techniques have been proposed in recent years. In this paper a basic survey on various NIR-VIS face matching techniques has been provided. These techniques categories into three parts namely invariant feature based, projection based and synthesis based. Methods in these categories address the modality gap between heterogeneous faces. Geometric and photometric normalization methods play an important role in reducing the cross modality gap and all existing methods achieve greater accuracy than commercial face matcher.

REFERENCES

[1] Lin, D., Tang, X.: Inter-modality face recognition. In: Leonardis, A. Bischof, H. Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 13–26. Springer, Heidelberg (2006)
[2] D. Yi, R. Liu, R.-F. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," in Proc. Int. Conf. Adv.Biometrics, 2007, pp. 523–530
[3] R. Wang, J.Yang, D. Yi, and S. Z. Li. An analysis-by synthesis method for heterogeneous face biometrics. In Proceedings of IAPR/IEEE International Conference on Biometrics,2009. 2
[4] Z. Lei and S. Li. Coupled Spectral Regression for matching heterogeneous faces. In zComputer Vision and Pattern Recognition, 2009. CVPR 2009.IEEE Conference on, pages 1123–1128, June 2009. 2
[5] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Li. Heterogeneous face recognition from local structures of normalized appearance. In Proc.3rd ICB, 2009.
[6] Jie Chen, Dong Yi, Jimei Yang, Guoying Zhao, S.Z. Li, and M. Pietikainen. 2009. Learning mappings for face synthesis from near infrared to visual light images. In CVPR. 156–163.
[7] S.Z. Li, Zhen Lei, and Meng Ao. 2009. The HFB Face Database for Heterogeneous Face Biometrics research.In CVPR. 1–8.
[8] B. Klare and A. Jain, "Heterogeneous Face Recognition: Matching NIR to Visible Light Images," Proc. Int'l Conf. Pattern Recognition,2010.
[9] Baochang Zhang, Lei Zhang, David Zhang, and Linlin Shen. 2010. Directional binary code with application to PolyU near-infrared face database. Pattern Recognition Letters (2010), 2337–2344.
[10] D. Goswami, Chi-Ho Chan, D. Windridge, and J. Kittler. 2011. Evaluation of face recognition system in heterogeneous environments (visible vs NIR). In ICCV. 2160–2167
[11] Likun Huang, Jiwen Lu, and Yap-Peng Tan. 2012. Learning modality-invariant features for heterogeneous

face recognition. In International Conference on Pattern Recognition (ICPR). 1683–1686.
[12] Z. Lei, C. Zhou, D. Yi, A. K. Jain, and S. Z. Li, “An improved coupled spectral regression for heterogeneous face recognition,” in Proc. 5th IAPR ICB, 2012, pp. 7–12.
[13] H. Maeng, S. Liao, D. Kang, S.-W. Lee, A .K. Jain, Nighttime face recognition at long distance: cross-distance and cross-spectral matching ,in: Proceedings of ACCV,2012,pp.708–721
[14] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li, “Coupled discriminant analysis for heterogeneous face recognition,” IEEE Trans. Inf. Forensics Security, vol. 7, no. 6, pp. 1707–1716, Dec. 2012.
[15] Xiong Pengfei, Lei Huang, and Changping Liu. 2012. A method for heterogeneous face image synthesis. In The IAPR International Conference on Biometrics (ICB). 1–6.
[16] Jun-Yong Zhu, Wei-Shi Zheng, and Jian-Huang Lai. 2013b. Logarithm Gradient Histogram: A general illumination invariant descriptor for face recognition. In Automatic Face and Gesture Recognition (FG).1–8.
[17] B. Klare and A. K. Jain,“Heterogeneous face recognition using kernel prototype similarities,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 6, pp. 1410–1422, Jun. 2013.
[18] S.Z. Li, Dong Yi, Zhen Lei, and Shengcai Liao. 2013. The CASIA NIR-VIS 2.0 Face Database. In CVPR. 348–353.
[19] Dihong Gong and Jiangyu Zheng. 2013. A Maximum Correlation Feature Descriptor for Heterogeneous Face Recognition. In Asian Conference on Pattern Recognition (ACPR). 135–139.
[20] Jun-Yong Zhu, Wei-Shi Zheng, Jian-Huang Lai, and S.Z. Li. 2014. Matching NIR Face to VIS Face Using Transduction. IEEE Transactions on Information Forensics and Security (2014), 501–514.
[21] Zhifeng Li, Senior Member, IEEE, Dihong Gong, Yu Qiao, Senior Member, IEEE, and zacheng Tao, Senior Member,” Common Feature Discriminant Analysis for Matching Infrared Face Images to Optical Face Images,” june 2014.