Near-IR to Visible Light Face Matching: Effectiveness of Pre-Processing Options for Commercial Matchers

John Bernhard jbernha3@nd.edu
Jeremiah Barr jbarr1@nd.edu
Kevin W. Bowyer kwb@nd.edu
Patrick Flynn flynn@nd.edu

Department of Computer Science and Engineering
University of Notre Dame, Notre Dame, IN 46556, USA

Abstract

The use of near-IR images for face recognition has been proposed as a means to address illumination issues that can hinder standard visible light face matching. However, most existing non-experimental databases contain visible light images. This makes the matching of near-IR face images to visible light face images an interesting and useful challenge. Image pre-processing techniques can potentially be used to help reduce the differences between near-IR and visible light images, with the goal of improving matching accuracy. We evaluate the use of several such techniques in combination with commercial matchers and show that simply extracting the red plane results in a comparable improvement in accuracy. In addition, we show that many of the pre-processing techniques hinder the ability of existing commercial matchers to extract templates. We also make available a new dataset called Near Infrared Visible Light Database (ND-NIVL) consisting of visible light and near-IR face images with accompanying baseline performance for several commercial matchers.

1. Introduction

Although the performance of face matchers has improved greatly over the years, dealing with the effects of varying illumination is a continuing difficulty. A method demonstrating some success with this problem is the use of Near Infrared (NIR) light images. By using NIR light to capture images, the lighting on the face can be controlled and is less affected by other lighting in the environment. NIR illumination is also used in surveillance for both regular and low-light situations. NIR face images captured by these surveillance systems would be matched against visible light images in existing face databases. One problem encountered when using NIR light images for face matching is that most existing face galleries consist exclusively of visible light images. This motivates the matching of faces images across different modalities. In the literature this is typically referred to as heterogeneous face matching. Although heterogeneous matching may refer to matching across other modalities, this paper will address NIR and visible light (VIS).

One method used to assist in matching NIR to VIS images is to perform pre-processing on images of both modalities in an attempt to transform them into images in which the illumination spectrum does not as overtly influence face appearance. Various pre-processing techniques have been proposed as effective means for matching NIR to VIS images. This paper is the first to demonstrate the effects of these common techniques on the performance of existing commercial matchers. Furthermore, this paper is the first to document the impact of extracting the red plane from the color images and using it in matching which results in consistently better accuracy with commercial matchers. This indicates that the commercial matcher performance levels used for baseline comparison in [4] and [11] are lower than necessary. Compared to previous papers in this area, this paper considers a larger number of pre-processing techniques as well as a larger number of commercial matchers. Lastly, this paper provides a new dataset of NIR and VIS images (ND-NIVL) as well as baseline performance results on this database for several existing commercial matchers.

2. Literature Review

There are three general categories of approach to heterogeneous face recognition: Face normalization, feature extraction, and subspace learning [6]. Papers on subspace learning include Lei et al. [6], Zhu et al. [17], Lin and Tang [10], and Lei and Li [5]. Those focusing on feature extraction include Yi et al. [16], Liao et al. [9], and Klare and Jain [4]. As we are using commercial matchers and do not have direct control over the feature extraction and learning, we focus on the face normalization step.

Various pre-processing techniques have been used to normalize the face images and reduce the differences in ap-

Paper	Datasets Used	Pre-Processing Techniques Used	Commercial Matchers Used
Lei et al., 2012 [6]	CASIA HFB extended	None	None
Zhu et al., 2012 [17]	CASIA HFB	None	None
Yi et al., 2009 [16]	MBGC 2008 portal challenge data	Laplacian of Gaussian filter	None
Lin and Tang, 2006 [10]	Own in-house data	None	None
Liao et al., 2009 [9]	CASIA HFB	Difference of Gaussian	None
Lei and Li, 2009 [5]	CASIA HFB	None	None
Klare and Jain, 2010 [4]	CASIA HFB	None	Cognitec FaceVACS
Goswami et al., 2011 [2]	Cross Spectral Dataset	Sequential chain retinex, self-quotient	None
Bourlai and Cukic, 2012 [1]	Own in-house data	Contrast limited adaptive histogram equalization, retinex, self-quotient, difference of Gaussian	None
Maeng et al., 2013 [11]	CASIA HFB, LDHF-DB	Histogram Equalization, Gaussian smoothing	Cognitec FaceVACS, PittPatt

Table 1: Papers using various techniques for heterogeneous face matching

pearance. Goswami et al. [2] use sequential chain, retinex and self-quotient pre-processing techniques. Bourlai and Cukic [1] use contrast limited adaptive histogram equalization, retinex, self-quotient, and difference of Gaussians filtering. Maeng et al. [11] use histogram equalization and Gaussian smoothing as pre-processing techniques, but the paper focuses on the effect of distance on heterogeneous face matching.

Table 1 shows the dataset, preprocessing techniques and commercial matchers used in each prior paper examined. The only papers found that give results for commercial matchers are [4], which compares the accuracy of its custom heterogeneous matcher with that of the Cognitec FaceVACS matcher, and [11], which compares against the accuracy of Cognitec FaceVACS and PittPatt. We found no papers that look at the effects of pre-processing techniques used in combination with existing commercial matchers. Also, we found no papers that consider simply extracting the red plane from the visible light images as a pre-processing technique. The motivation for using only the red plane is that, of the RGB planes, it is closest to the near-IR wavelength. Thus, in that sense, the red plane image should have greater similarity to a near-IR image than does the RGB or grayscale image.

Only a few NIR/VIS face databases have been made available to the research community. The CASIA HFB [7] database was made available in 2009 with 100 subjects and then expanded in 2010 to 5097 images of 202 subjects. In 2013, a new CASIA database was released to replace the HFB database and facilitate future research [8]. This database consists of 17580 images of 725 subjects. Of the 725 subjects, 14 appear in multiple sessions. Each image in the database is 640x480 pixels. The images contain varia-

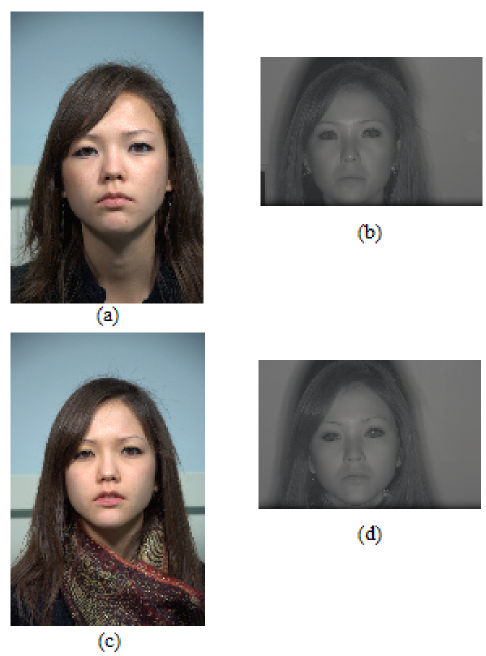

Figure 1: Example images from the ND-NIVL. (a) Visible light image taken fall 2011. (b) Near-IR light image taken fall 2011. (c) Visible light image taken spring 2012. (d) Near-IR light image take spring 2012.

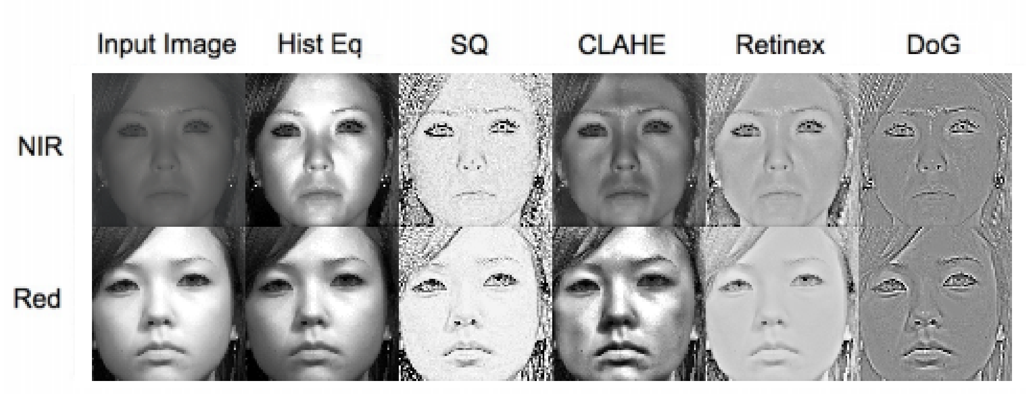

Figure 2: Examples of pre-processing techniques used

tions in pose, expression, eyewear, and distance. The CASIA HFB database is (at the time of writing) the most commonly used database for the heterogeneous NIR to VIS face matching problem.

Another available NIR/VIS face database is the LDHF-DB [11]. This database has 1600 images of 100 subjects. All images were taken during a single session. The images contain variations in distance ranging from 1m to 150m. The images have a resolution of 5,184x3,456.

The Cross Spectral Dataset [2] consists of 4189 images of 430 subjects and contains variation of pose. Although the paper states that the dataset is available to the research community, we were informed by the authors that it is not being distributed.

Our ND-NIVL dataset is the largest available dataset with 24605 images of 574 subjects. Of the 574 subjects 402 appear in multiple sessions. The NIR images have a resolution of 4770x3177 and the visible light images have a resolution of 4288x2848. This makes ND-NIVL the largest database of high-resolution NIR and VIS images.

3. Dataset and Experimental Method

The dataset used for this paper (ND-NIVL) was collected over the course of two semesters (fall 2011 and spring 2012). All data were collected under the terms of a protocol approved by the institution's human subjects institutional review board. The VIS images were collected using a Nikon D90 camera. The Nikon D90 uses a 23.6x15.8 mm CMOS sensor and the resulting images have a 4288x2848 resolution. The images were acquired using automatic exposure and automatic focus settings. All images were acquired under normal indoor lighting at about a 5-foot standoff with frontal pose and a neutral facial expression.

The NIR images were acquired using a Honeywell CFAIRS system. CFAIRS uses a modified Canon EOS 50D camera with a 22.3x14.9 CMOS sensor. The resulting images have a resolution of 4770x3177. All images were acquired under normal indoor lighting with frontal pose and neutral facial expression. NIR images were acquired at both a 5ft and 7ft standoff.

The dataset contains a total of 574 subjects. There are a total of 2,341 VIS images and 22,264 NIR images from the 574 subjects. A total of 402 subjects had both VIS and NIR images acquired during at least one session during both the fall and spring semesters. Both VIS and NIR images were acquired in the same session, although not simultaneously. One VIS image was acquired per subject per session. Around 10 NIR images were acquired per subject per session. Figure 1 shows example images from the dataset. For the experiments in this paper, the NIR images were resized to 720x480 and the VIS images were resized to 480x721. The inter-ocular distance still varies between images. For each commercial matcher, two matching experiments were performed with various pre-processing techniques used before each. The first experiment matched NIR images from the spring 2012 acquisitions against the VIS images from fall 2011. The second experiment matched NIR images from the fall 2011 acquisitions against the VIS images from the spring 2012 acquisitions.

The following pre-processing techniques were performed on the cropped face images prior to template extraction and matching. Then INface toolbox v2.0 for MATLAB [14][13] was used for each pre-processing technique after extracting the red channel. The default parameters were used for each function. Figure 2 shows examples of each.

- Red - the red channel is extracted from the VIS image. The motivation behind this technique is to select the color channel closest in wavelength to infrared. We know of no previous work that looked at the performance of commercial or custom-built matchers using the red plane as the visible light image.

- Histogram Equalization - Histogram equalization is applied to the red channel of the VIS image as well as to the NIR image. The motivation behind this is to normalize the images toward the same contrast distribution. This pre-processing step was investigated in [11] for use with their custom-built matcher.

- CLAHE - Contrast Limited Adaptive Histogram Equalization is applied to the red channel of the VIS image as well as to the NIR image. The motivation here is to flatten the intensity values independently for different regions of the image [12]. This pre-processing step was investigated in [1] for use with their custom-built matcher.

- SQ - Self-Quotient is applied to the red channel of the VIS image as well as to the NIR image. Self-quotient normalizes each pixel by the average value in the region [15]. This pre-processing step was explored in [2] for use with their custom-built matcher.

- Retinex - The Retinex operation is applied to the red channel of the VIS image as well as to the NIR image. The Retinex operation compresses the range of intensity values of an image [3]. This pre-processing step was examined in [2] for use with their custom-built matcher.

- DoG - The Difference of Gaussian filter is applied to the red channel of the VIS image as well as to the NIR image. DoG filters out low frequency components of an image. This pre-processing step was investigated in [9] for use with their custom-built matcher.

4. Baseline Performance of Commercial Matchers

Of the 24,605 images in the ND-NIVL dataset, commercial matcher #2 was able to create a template 96.74% of the time. Commercial matcher #1 was able to create a template for 97.49% of the images, and Cognitec FaceVACS performed the best with a template yield of 99.97%. Figure 3 shows the resulting ROC curve for NIR to VIS matching and NIR to red plane matching for each commercial matcher for the first experiment; Figure 4 shows the same for the second experiment. In both cases it can be seen that matching NIR images against the red plane VIS images offers an improvement over simply matching the NIR images against the raw VIS images for each commercial matcher. The figures also show that Cognitec FaceVACS significantly outperforms the other two commercial matchers on this dataset.

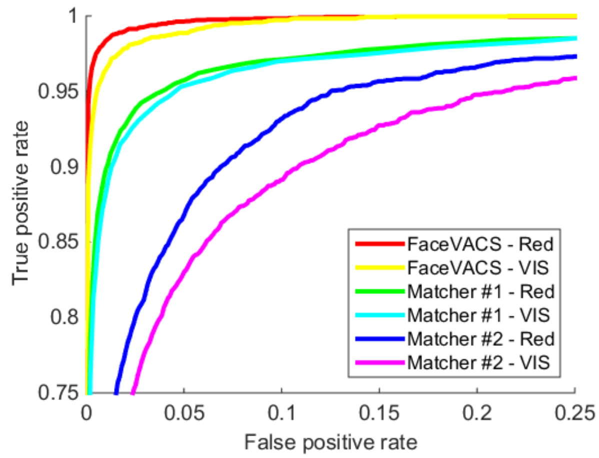

Figure 3: 2011 NIR vs 2012 VIS and red plane image ROC curves

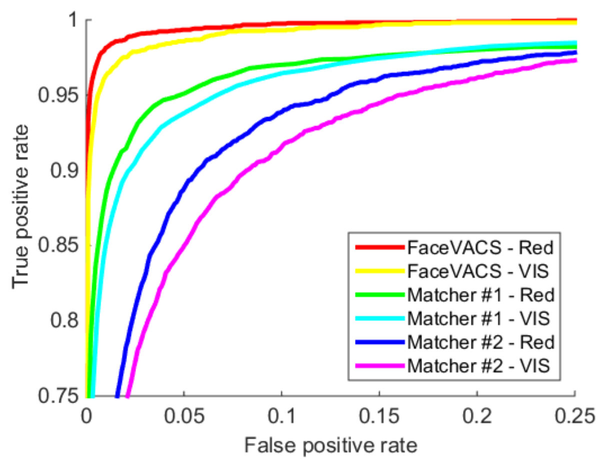

Figure 4: 2012 NIR vs 2011 VIS and red plane image ROC curves

5. Evaluation of Pre-Processing Operations

In addition to the red plane pre-processing, the other pre-processing techniques mentioned in section 3 were also used. These techniques have been shown to be useful with matchers developed in-house. However, when applied prior to use of commercial matchers there is a cost in the number of templates that are able to be generated.

Table 2 shows the number of templates that were generated by each commercial matcher under each pre-processing technique. As can be seen in the table, Cognitec FaceVACS is able to extract templates without issue for most of the pre-processing techniques with the exception of Retinex and DoG. However, commercial matcher

Matcher	Red	Retinex	DoG	Histeq	SQ	CLAHE
Commercial matcher #1	97.49%	63.45%	1.19%	0%	0%	61.50%
Commercial matcher #1*	-	63.91%	1.07%	67.30%	92.67%	62.40%
Commercial matcher #2	96.74%	11.32%	0.56%	91.52%	67.12%	66.33%
Commercial matcher #2*	-	6.15%	0.08%	73.93%	69.20%	53.98%
Cognitec FaceVACS	99.97%	86.26%	11.78%	99.95%	99.86%	99.62%

Table 2: Percent yield for template extraction for each commercial for various pre-processing technique
* Indicates that the pre-processing was done before the faces were cropped from the original image instead of after

#1 and commercial matcher #2's template yield is significantly impacted by all techniques other than red plane. To see if the template yield for commercial matchers #1 and #2 could be increased, we repeated the experiments with the pre-processing techniques applied before the face images were cropped from the original images instead of afterward. This improved the yield in some instances, but decreased it in others. The only pre-processing technique to yield a reasonable number of templates for each commercial matcher is red plane extraction.

Figures 5 and 6 show the ROC curves using each pre-processing technique for the most accurate commercial matcher. Since the number of templates extracted for each pre-processing technique differs, these ROC curves are not directly comparable. The ROC curve for retinex pre-processing appears to have high accuracy, but since a template was only created for 86.26% of the face images, it should not be compared to those pre-processing techniques that have over 99% template yield. Of the other techniques, red plane, histogram equalization and self quotient were the highest performing and all resulted in an improvement over matching the NIR images to the raw VIS images.

Results were also obtained for Cognitec FaceVACS where the pre-processing was applied prior to the faces being cropped from the image. Figure 7 shows said results. Only the pre-processing techniques that provided high template yields are shown. Again, we can see that red plane performs best. The other techniques perform similarly to when pre-processing is performed after the face is cropped from the image with the exception of histogram equalization. In this case, when histogram equalization is applied to the red plane of the whole image instead of just the cropped face, the faces tend to be washed out. This explains the relativity poor performance of histogram equalization in this case.

In order to see if the results could be improved further, fusion was performed across commercial matchers and across pre-processing techniques. As all commercial matchers tested performed well with red plane images, fusion of those results were performed. Figure 8 shows mean score fusion of the three commercial matchers using the red plane images. All scores were min-max normalized before

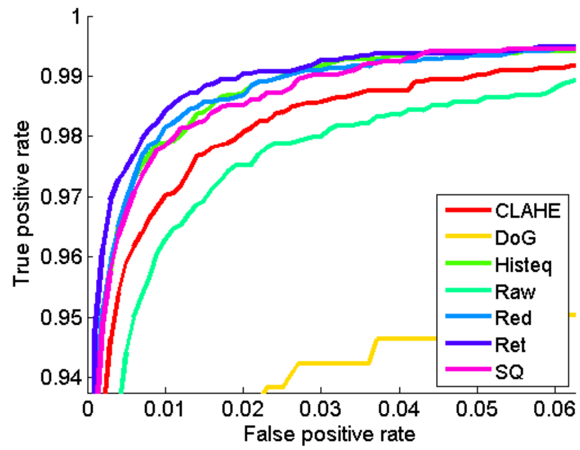

Figure 5: 2011 NIR vs 2012 pre-processed VIS image ROC curves for Cognitec FaceVACS. Ret and DoG should not be directly compared to others due to low template yield.

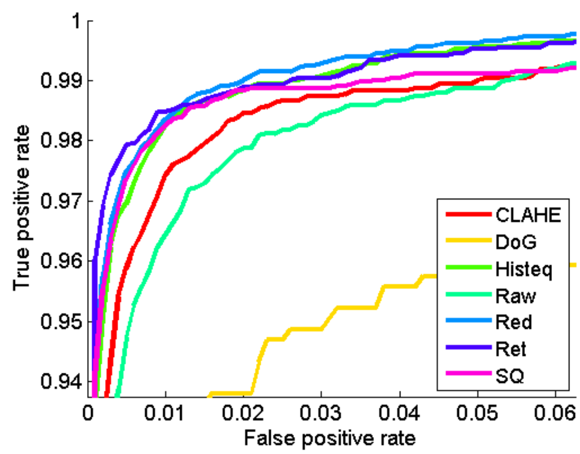

Figure 6: 2012 NIR vs 2011 pre-processed VIS image ROC curves for Cognitec FaceVACS. Ret and DoG should not be directly compared to others due to low template yield.

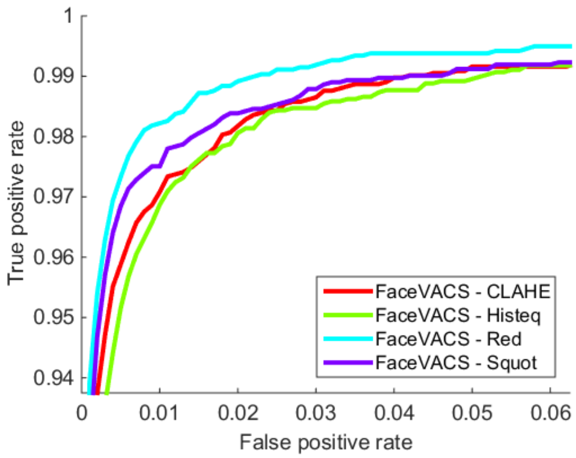

Figure 7: 2012 NIR vs 2011 pre-processed VIS image ROC curves for Cognitec FaceVACS where pre-processing was performed prior to the faces being cropped

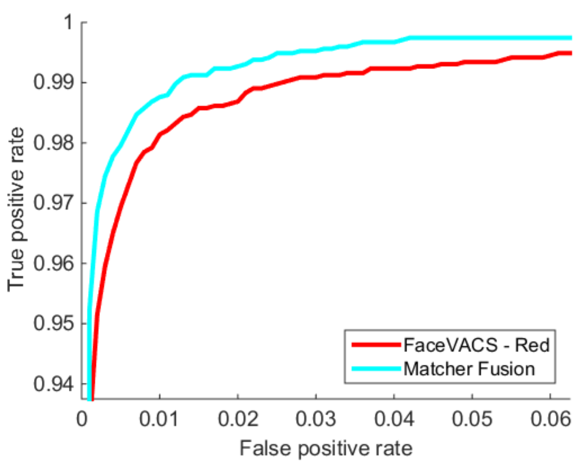

Figure 8: ROC curves comparing mean score fusion of the three commercial matchers to Cognitec FaceVACS

fusion. As can be seen in figure 8, there is an improvement in the resulting ROC curve when fusing the results of the three matchers.

Fusion across pre-processing techniques was performed using the Cognitec FaceVACS matcher and the red, histogram equalization, self quotient, and CLAHE pre-processing techniques. The other pre-processing techniques were not considered due to poor yield for template extraction. Figure 9 shows the results of max score fusion of the four pre-processing techniques compared to the results using only red plane images. Again, the fusion provides a slight improvement over the non-fusion results.

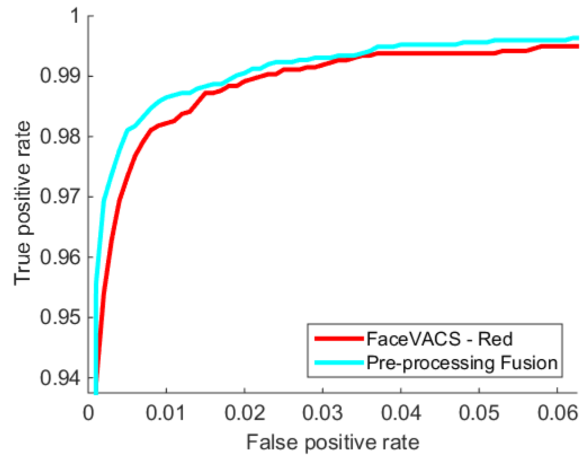

Figure 9: ROC curves comparing max score fusion of four pre-processing techniques to red plane only

6. Conclusions

This paper is the first to show the effects of various pre-processing techniques on the results from commercial matchers when performing heterogeneous NIR to VIS matching. We have shown that comparing NIR images to red plane VIS images gives better results than comparing NIR images to VIS images when using commercial matchers, and that the resulting improvement in accuracy is comparable to that of other common pre-processing techniques. Previous work on NIR-VIS face matching has not evaluated the effect of using the red plane from visible light images versus using the full color images. Furthermore, pre-processing techniques can significantly impact the ability of commercial matchers to produce templates, and the degree to which this is the case depends greatly on when the pre-processing is performed. Of the commonly used pre-processing techniques, only red plane produces a reasonable number of templates for each commercial matcher.

Given these results, this paper establishes a rigorous baseline against which future custom NIR to VIS matchers should compare themselves. Future NIR to VIS matching techniques should be compared to the results of matching NIR against red plane VIS images with Cognitec FaceVACS or an equally high-performing commercial matcher in order to demonstrate a practical improvement.

We have also provided a baseline for heterogeneous NIR to VIS matching on multiple commercial matchers on a newly available dataset, and have shown that Cognitec FaceVACS outperforms commercial matcher #1 and commercial matcher #2 on this dataset. This has only been done before with two commercial matchers by Maeng et al. [11].

Lastly, multiple previous papers have described NIR-

VIS databases, but many are not available [10][2][1], have single session data only [11], or have lower resolution images [8] than the dataset we provide. ND-NIVL as described in section 3 is available and has in fact already been distributed to one other research group. For details on obtaining the ND-NIVL dataset please refer to the following website: http://www3.nd.edu/ cvrl/CVRL/Data_Sets.html.

Acknowledgements

This applied research effort was supported by the U.S. government.

References

[1] T. Bourlai and B. Cukic. Multi-spectral face recognition: identification of people in difficult environments. In *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*, pages 196–201. IEEE, 2012.

[2] D. Goswami, C. H. Chan, D. Windridge, and J. Kittler. Evaluation of face recognition system in heterogeneous environments (visible vs nir). In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, pages 2160–2167. IEEE, 2011.

[3] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997.

[4] B. Klare and A. K. Jain. Heterogeneous face recognition: Matching nir to visible light images. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1513–1516. IEEE, 2010.

[5] Z. Lei and S. Z. Li. Coupled spectral regression for matching heterogeneous faces. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1123–1128. IEEE, 2009.

[6] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li. Coupled discriminant analysis for heterogeneous face recognition. *IEEE Transactions on Information Forensics and Security*, 7(6):1707–1716, 2012.

[7] S. Z. Li, Z. Lei, and M. Ao. The hfb face database for heterogeneous face biometrics research. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 1–8. IEEE, 2009.

[8] S. Z. Li, D. Yi, Z. Lei, and S. Liao. The casia nir-vis 2.0 face database. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 348–353. IEEE, 2013.

[9] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li. Heterogeneous face recognition from local structures of normalized appearance. In *Advances in Biometrics*, pages 209–218. Springer, 2009.

[10] D. Lin and X. Tang. Inter-modality face recognition. In *Computer Vision–ECCV 2006*, pages 13–26. Springer, 2006.

[11] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain. Nighttime face recognition at long distance: Cross-distance and cross-spectral matching. In *Computer Vision–ACCV 2012*, pages 708–721. Springer, 2013.

[12] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

[13] V. Štruc and N. Pavešić. Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica*, 20(1):115–138, 2009.

[14] V. Štruc and N. Pavešic. Photometric normalization techniques for illumination invariance. *Advances in Face Image Analysis: Techniques and Technologies, IGI Global*, pages 279–300, 2011.

[15] H. Wang, S. Z. Li, Y. Wang, and J. Zhang. Self quotient image for face recognition. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 2, pages 1397–1400. IEEE, 2004.

[16] D. Yi, S. Liao, Z. Lei, J. Sang, and S. Z. Li. Partial face matching between near infrared and visual images in mbgc portal challenge. In *Advances in Biometrics*, pages 733–742. Springer, 2009.

[17] J.-Y. Zhu, W.-S. Zheng, and J. Lai. Transductive vis-nir face matching. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 1437–1440. IEEE, 2012.