An Analysis-by-Synthesis Method for Heterogeneous Face Biometrics

Rui Wang, Jimei Yang, Dong Yi, and Stan Z. Li*

Center for Biometrics and Security Research,
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
{rwang,jmyang,dyi,szli}@cbsr.ia.ac.cn

Abstract. Face images captured in different spectral bands, *e.g.*, in visual (VIS) and near infrared (NIR), are said to be heterogeneous. Although a person's face looks different in heterogeneous images, it should be classified as being from the same individual. In this paper, we present a new method, called *face analogy*, in the analysis-by-synthesis framework, for heterogeneous face mapping, that is, transforming face images from one type to another, and thereby performing heterogeneous face matching. Experiments show promising results.

Keywords: Heterogenous face biometrics, face analogy, face matching, analysis-by-synthesis.

1 Introduction

Different types face biometrics have been developed, including visual (VIS) [1], near infrared (NIR) [2], thermal infrared (TIR) [3], and 3D [4] image based. In any of these methods, it is assumed that both enrollment and query face images are of the same type. Face matching two different image types are said to be *heterogeneous* because the image formation characteristics are different.

Although heterogeneous face images of a given person have different face appearances by pixel values, the identify of the face images should be classified as the same whichever type of image the face is contained in. Face biometrics by matching between heterogeneous face images are collectively called *heterogeneous face biometrics* (HFBs).

Recent developments have led to several proposals of HFBs, that between VIS and face sketch [5], between VIS and NIR [6], and between 3D and NIR [7], and also reconstruction of the facial shape from an NIR image [8]. The MBGC (Multiple Biometric Grand Challenge) tests organized by NIST (National Institute of Standards and Technology) has devised the modality of matching between VIS and NIR face images as one of its experiments [9], with the purpose to examine how well NIR-VIS face biometric could be done and how its fusion with VIS-VIS face biometric could improve the overall performance.

In this paper, we present a new method, called *face analogy*, for HFBs, in the analysis-by-synthesis framework. The face analogy method transforms face images

---

* Corresponding author.

from one type to another. Heterogeneous face matching can thereby be performed. Experiments show promising results.

2   Image Modeling

2.1   Spectral Image Formation

Images used in face recognition is related to facial shape, skin and hair. A 3D face image is related to the shape only. It is captured by a range measuring system usually made from a laser range system or stereo vision system. Represent a range image taken from a viewpoint by $z(x,y)$. The pixel values measure the distances from the sensor to the facial surface points.

Spectral face images are different from 3D in nature. Fig. 1 illustrates ultra-violet (UV), VIS and infrared (IR) spectral bands in different wavelengths. The IR portion of electromagnetic wave can be divided into four spectral regions: near infrared (NIR), short-wave infrared, thermal infrared (TIR), and far infrared (FIR).

Ultraviolet	Visible		Infrared		
		Near IR	Short-wave IR	Thermal IR	Far IR

| 0.4 μm | 0.7 μm | 1.1 μm | 2.4 μm | 15 μm |

Fig. 1. Spectral bands of electromagnetic wave radiation

The Lambertian law provides an image formation model, relating a spectral image with the 3D shape of the sensed object, the object surface properties and the illumination source:

$$I(x,y) = \rho(x,y)\mathbf{n}(x,y)\mathbf{s} \tag{1}$$

where $I(x,y)$ is the spectral image, $\rho(x,y)$ is the albedo of the facial surface material at point (x,y) (also changing with wavelengths), $\mathbf{n} = (n_x, n_y, n_z)$ is the surface normal (a unit row vector) at the 3D surface point $z(x,y)$, and $\mathbf{s} = (s_x, s_y, s_z)$ is the point lighting direction (a column vector, with magnitude). The normal directions $\mathbf{n}(x,y)$ may be derived from the range image $z(x,y)$ but not vice versa.

2.2   Face Skin Spectroscopy

HFBs across spectral bands are related to facial skin and hair spectroscopy. Skin spectropy is a field of study on the interaction between electromagnetic wave and human skin. Essentially, skin spectroscopy studies the albedo ρ of skin as a function of wavelength. Understanding face spectroscopy helps formulating heterogeneous face image formation models and designing HFB algorithms.

The optics of skin is dependent on wavelength and dose of the incident light. These are the main bands that have been used practically for for non-invasive biometric applications. The spectrum of VIS is usually divided into three channels: red, green and blue.

In the VIS spectrum, melanin is the only pigment affecting the transmittance of epidermis. Shorter wavelengths, such as blue light, are highly absorbed by certain tissue components such as melanin and blood relative to longer wavelengths. In addition, the optical scattering increases as wavelength getting shorter in this spectral range.

The skin color is affected by the type and quantity of melanin in epidermis. This results in various skin colors from white to brown to black though the structure of human skin is similar across different race. The absorption of melanin decreases monotonously from the short wavelength of 0.25μm in the UV band (through the VIS band wavelength) to the long wavelength of 1.1μm in the NIR band.

Beyond 1.1μm, the absorption of melanin is negligible. Both transmittance and remittance of skin are uncorrelated with melanin and the most of IR is transmitted through epidermis into dermis. TIR of wavelength between 2.4 to 15 μm is often used passively to determine the temperature of skin. The amount of heat radiation emitted by skin increases with temperature, therefore thermography allows one to see variations in temperature.

The dermis has significantly different structure and composition with epidermis, In the dermis, the scattering effect dominates. Many observations show the transmittance and remittance are close to 100% across the spectrum from 0.3μm-2.4μm, indicating that very little radiation is absorbed there. Light at longer wavelengths can penetrate the dermis deeper than shorter wavelengths.

3   The Face Analogy Method

In the analysis-by-synthesis approach to HFBs, face analogy is a method for converting face images from one type to another by transforming facial textures. It shares the idea of "image analogies" [10] from the computer graphics community. For the NIR-VIS face analogy, an NIR face can be transformed analogously into a VIS face as if the lighting changes, based on a pair of well-registered NIR-VIS faces.

The NIR to VIS transform also has a practical motivation: We would like to take NIR images as the probe, take advantage that active NIR imaging can tackle the problem of variable illumination, and match against VIS faces of target.

Image analogy has two assumptions. One is that in any location of image, one type of texture pattern and the other type of texture pattern always co-occur. The other is that in any given training image pair, there are sufficient sorts of co-occurrent texture patterns so that only one pair of registered images are needed as training data.

However, these two assumptions tend to fail when applied to faces. Facial texture patterns are usually localized and thus co-occurrence just locally works. It means that texture pattern co-occurrence can only be trusted in the same facial semantic region, such as nose, eyes and corners of mouth. For example, if in an NIR image a texture pattern in the nose area is similar to one in the mouth area, their corresponding VIS texture patterns in the same areas cannot be guaranteed being similar.

We solve this problem in face analogy by introducing more pairs of well-registered NIR-VIS faces. NIR-VIS texture patterns coming from the same region of different face pairs together compose a training database for this region. Consequently, texture patterns in this region of a new NIR face can be found in the corresponding training database and then this NIR texture patterns can be transformed to VIS texture pattern by analogy. Further, by the analysis of NIR and VIS imaging models, we establish a common invariant using local normalization technique [11] and thereby parameterize the synthesis procedure.

As a result, the procedure of face analogy equivalently consists of a de-lighting (local normalization) step and a re-lighting (synthesis) step. In the following, we analyze VIS and NIR imaging models and then provide a procedure for face analogy.

3.1 Local Image Models

Assuming that the facial surface is Lambertian, then like Eq.1 a VIS face imaging model under spot lighting source is modeled by

$$I'(x,y) = \rho'(x,y)\mathbf{n}(x,y) \cdot \mathbf{s}' \tag{2}$$

When NIR lighting is near spot source, we have

$$I(x,y) = \rho(x,y)\frac{\mathbf{n}(x,y) \cdot \mathbf{s}}{r^2(x,y)} \tag{3}$$

where $r(x,y)$ is the distance from spot lighting to face. Note that from spectroscopy, the albedo functions $\rho'(x,y)$ for VIS and $\rho(x,y)$ for NIR are different. These two equations will be abbreviated as $I'(i) = \rho'(i)\mathbf{n}(i) \cdot \mathbf{s}'$ and $I'(i) = \rho'(i)\frac{\mathbf{n}(i) \cdot \mathbf{s}}{r^2(i)}$.

Without loss of generality, we assume that facial surface normal and lighting vary continuously, while more accurate formulation would take into account abrupt changes near boundaries. Then, in a small neighborhood $\mathcal{N}_i$ of pixel i, we have $\mathbf{n}(i) \cdot \mathbf{s}' = c'$ and $\mathbf{n}(i) \cdot \mathbf{s}/r^2(i) = c$, where c' and c are some constants, and the Lambertians become,

$$I'(i) = \rho'(i)c', i \in \mathcal{N}_i, \tag{4}$$
$$I(i) = \rho(i)c, i \in \mathcal{N}_i \tag{5}$$

3.2 Local Normalization

To construct common local invariants of NIR and VIS face images, we introduce an operator called *local normalization* as

$$f: h(i) \leftarrow \frac{I(i) - m(\mathcal{N}_i)}{\sigma(\mathcal{N}_i)}, \tag{6}$$

where $m(\mathcal{N}_i)$ and $\sigma(\mathcal{N}_i)$ denote the mean and the deviation within $\mathcal{N}_i$, respectively. Similarly define

$$g: h'(i) \leftarrow \frac{I'(i) - m'(\mathcal{N}_i)}{\sigma'(\mathcal{N}_i)}, \tag{7}$$

Taking the Lambertian equations into consideration, we get

$$h(i) = \frac{I(i) - m(\mathcal{N}_i)}{\sigma(\mathcal{N}_i)} = \frac{\rho(i) - \bar{\rho}(\mathcal{N}_i)}{\sqrt{\sum(\rho(i) - \bar{\rho}(\mathcal{N}_i))^2/n^2}}, \tag{8}$$

$$h'(i) = \frac{I'(i) - m'(\mathcal{N}_i)}{\sigma'(\mathcal{N}_i)} = \frac{\rho'(i) - \bar{\rho}'(\mathcal{N}_i)}{\sqrt{\sum(\rho'(i) - \bar{\rho}'(\mathcal{N}_i))^2/n^2}}, \tag{9}$$

Both $h(i)$ and $h'(i)$ are functions of the albedos.

We would like to model the relationship of $h(i)$ and $h'(i)$. Assume that there exists a local linear transformation $T(\mathcal{N}_i)$ from $\rho(i)$ to $\rho'(i)$, that is,

$$\rho'(i) = T(\mathcal{N}_i) \circ \rho(i), i \in \mathcal{N}_i. \tag{10}$$

Then, we can use the local linear transformations to equate

$$h'(i) = h(i) \tag{11}$$

3.3   Procedure

According to the analysis above, it is clear that we will accomplish the analogy from I to I' by applying a transformation $g^{-1} \circ f$, that is,

$$I' = g^{-1} \circ f \circ I, \tag{12}$$

where g^{-1} is,

$$g^{-1}: I'(i) \leftarrow h'(i) * \sigma'(\mathcal{N}_i) + m'(\mathcal{N}_i), \tag{13}$$

More specifically

$$I'(i) = \frac{I(i) - m(\mathcal{N}_i)}{\sigma(\mathcal{N}_i)} \cdot \sigma'(\mathcal{N}_i) + m'(\mathcal{N}_i). \tag{14}$$

Therefore, we can achieve face analogy when we can estimate these parameters.

Given a set of NIR faces $S = \{I_k\}$ and a set of VIS faces $S' = \{I'_k\}$, suppose that for each pair k, I_k and I'_k are mutually registered. Face Analogy takes S, S' and a new NIR face I as the input, and synthesizes the VIS counterpart I' of I as

$$I' \leftarrow \text{FaceAnalogy}(S, S', I). \tag{15}$$

The process of synthesizing a VIS face I' from an NIR face I is pixel-wise and in scan-line order. For each $i \in I'$, we first estimate $m(\mathcal{N}_i)$ and $\sigma(\mathcal{N}_i)$ by Eq.9, then search for the best match $I_k(\mathcal{N}_i)$ of $I(\mathcal{N}_i)$ among all NIR facial patterns defined in $\mathcal{N}_i$ and finally obtain the corresponding VIS facial pattern as

$$I'_k(\mathcal{N}_i) \leftarrow \text{BestMatch}(S, S', I, I', i). \tag{16}$$

So, we can estimate $\hat{m}'(\mathcal{N}_i)$ and $\hat{\sigma}'(\mathcal{N}_i)$ from $I'_k(\mathcal{N}_i)$ by Equ.13, and further obtain $I'(i)$ by Equ.(14). To reduce the error induced by scan-line, I' can be synthesized in four directions: from upper-left to bottom-right, bottom-right to upper-left, lower-left to upper-right, and upper-right to lower-left, and then their mean is computed as the final result.

4   Experiments

The following experiments are done to illustrate how to apply face analogy to synthesize a VIS face from a NIR probe face in terms of target VIS faces, and how to match the synthesized VIS face to the target face.

A database of NIR and VIS face image pairs of 100 persons was collected for the training and testing of face analogy and heterogeneous face matching. There are two NIR images and two VIS face images for each person, and so there are 200 NIR images and 200 VIS images in the database. The VIS and NIR images are captured in different time, containing variations in pose, expression, light condition, and so on. The face images are aligned with 68 landmarks manually into the size of 128×128 pixels. Fig.2 show some face images after alignment.
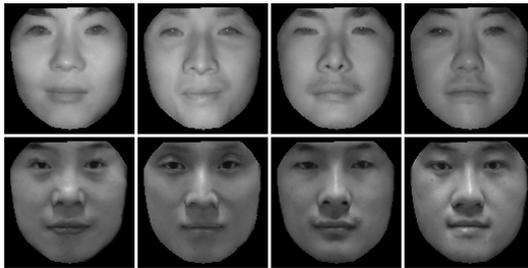

Fig. 2. Examples of aligned NIR (upper) and VIS (lower) face images

The 200 images in the NIR set are used as the source images from which the VIS (SYN) images are synthesized. The synthesis by face analogy is done in the leave-one-out manner, that is, by using all except one VIS images as the references. Fig. 3 shows some face analogy synthesized results. From the results, we can see the face analogy algorithm performs reasonably well for the NIR-VIS synthesis problem. The structure of NIR face is reserved meanwhile the appearance is transformed to the VIS type.

For the SYN-VIS matching, the 200 images in the NIR set are used as the probes, and the 200 images in the VIS set are used as the targets. No training is needed. The NIR probes are transformed to the counterpart VIS images (SYN) using face analogy. The SYN faces are then matched against the target of the VIS faces in the test set. Four matching methods are evaluated:

  – NIR-VIS: face matching using the NIR-VIS image data directly.
  – SYN-VIS: face matching using the SYN-VIS image data directly.
  – NIR-VIS (LoG): face matching using Laplacian of Gaussian (LoG) filtered image data.
  – SYN-VIS (LoG): face matching using LoG filtered image data.

The ROC curves of the four methods are shown in Fig. 4. In this experiment, the SYN-VIS methods outperform the rudimental NIR-VIS methods remarkably. Filtering NIR and VIS face images by LoG can further improve the performance.

Fig. 3. Face analogy synthesized results: Source NIR (upper), synthesized VIS (middle), and true VIS (lower) face images
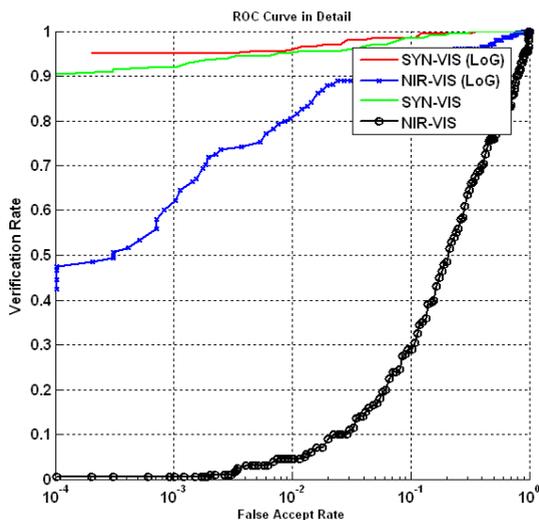

Fig. 4. The ROC curves of NIR-VIS heterogenous matchings and SYN-VIS matchings based on face analogy

5   Conclusions

Heterogeneous face biometrics (HFBs) perform biometric matching across heterogeneous face images. We have provided an analysis of problems in HFBs, and present a new method, face analogy, for transforming one type of face image to another, and use the synthesized image as a basis for NIR-VIS matching. Experiments are provided to illustrate the analysis-by-synthesis based algorithm.

Although the present analysis-by-synthesis method is for cross-spectrum face mapping between NIR and VIS types, cross-modality mapping such as between VIS and 3D

could be tackled in a similar way. HFBs could be used standalone for biometric authentication or work as an added module to improve face recognition using homogeneous face images. Promising results are obtained.

References

1. Zhao, W., Chellappa, R., Phillips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys, 399–458 (2003)
2. Li, S.Z., Chu, R., Liao, S., Zhang., L.: Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (2007) (Special issue on Biometrics: Progress and Directions)
3. Kong, S.G., Heo, J., Abidi, B., Paik, J., Abidi, M.: Recent advances in visual and infrared face recognition - A review. Computer Vision and Image Understanding 97(1), 103–135 (2005)
4. Bowyer, K.W., Chang, Flynn, P.J.: A survey of 3D and multi-modal 3D+2D face recognition. In: Proceedings of International Conference on Pattern Recognition, pp. 358–361 (2004)
5. Tang, X., Wang, X.: Face sketch recognition. IEEE Transactions on Circuits and Systems for Video Technology 14(1), 50–57 (2004)
6. Yi, D., Liu, R., Chu, R., Lei, Z., Li, S.Z.: Face matching between near infrared and visible light images. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 523–530. Springer, Heidelberg (2007)
7. Yang, W., Yi, D., Lei, Z., Sang, J., Li, S.Z.: 2D-3D face matching using CCA. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition (2008)
8. Lei, Z., Bai, Q., He, R., Li, S.Z.: Face shape recovery from a single image using cca mapping between tensor spaces. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2008)
9. NIST: Multiple Biometric Grand Challenge (MBGC) (2008), http://face.nist.gov/mbgc
10. Hertzmann, A., Jacobs, C.E., Oliver, N., Curless, B., Salesin, D.H.: Image analogies. In: Fiume, E. (ed.) SIGGRAPH 2001, Computer Graphics Proceedings, pp. 327–340. ACM Press / ACM SIGGRAPH (2001)
11. Xie, X., Lam, K.M.: An efficient illumination normalization method for face recognition. Pattern Recognition Letters 27, 609–617 (2006)