HETEROGENEOUS FACE RECOGNITION: AN EMERGING TOPIC IN BIOMETRICS

Contributor

Guodong Guo
West Virginia University

An emerging topic in biometrics is matching between heterogeneous image modalities, called heterogeneous face recognition (HFR). This emerging topic is motivated by the advances in sensor technology development that make it possible to acquire face images from diverse imaging sensors, such as the near infrared (NIR), thermal infrared (IR), and three-dimensional (3D) depth cameras. It is also motivated by the demand from real applications. For example, when a subject's face can only be acquired at night, the NIR or IR imaging might be the only modality for acquiring a useful face image of the subject. Another example is that no imaging system was available to capture the face image of a suspect during a criminal act. In this case a forensic sketch, drawn by a police artist based on a verbal description provided by a witness or the victim, is likely to be the only available source of a face of the suspect. Using the sketch to search a large database of mug-shot face photos is also a heterogeneous face recognition problem. Thus it is interesting to study the HFR as a relatively new topic in biometrics. In this article, several specific HFR problems are presented, and various approaches are described to address the heterogeneous face matching problems. Some future research directions are discussed as well to advance the research on this emerging topic.

Introduction

Biometrics is about the identification of humans by their characteristics or traits, which include both physiological and behavioral characteristics. Physiological traits are related to the body shape, such as face, fingerprint, and iris, while behavioral characteristics are related to the pattern of human behavior, such as the typing rhythm, gait, and voice.

"Biometrics is about the identification of humans by their characteristics or traits, which include both physiological and behavioral characteristics."

Because of the important and useful applications, such as identity management, law enforcement, and surveillance, biometrics has been an active research topic in the field of computer vision and pattern recognition.

Among various biometric traits, face recognition is one of the most challenging research topics, since there are many possible variations that affect the face matching performance. In traditional face recognition studies, the focus has been on addressing the changes and variations caused by human aging, head pose, illumination, and facial expressions, called A-PIE. Although significant progresses have been made especially for addressing the PIE problems, new challenges are emerging.

One of the emerging topics in face biometrics is matching between heterogeneous image modalities, called heterogeneous face recognition (HFR). This emerging topic is motivated by the advances in sensor technology development that make it possible to acquire face images from diverse imaging sensors, such as the near infrared (NIR), thermal infrared (IR), and three-dimensional (3D) depth cameras. It is also motivated by the demand from real applications. For example, when a subject's face can only be acquired at night, the NIR or IR imaging might be the only modality for acquiring a useful face image of the subject. Thus it is interesting to study the HFR as a relatively new topic in biometrics.

In this article, several specific problems belonging to HFR will be presented in the section "Heterogeneous Face Recognition Problems," and different HFR algorithms and approaches will be introduced in the section "Heterogeneous Face Recognition Algorithms." Various HFR databases will be described briefly in the section "Heterogeneous Face Databases." Future research directions for HFR are discussed in the section "Some Thoughts on Future Directions. This is followed by "Concluding Remarks."

Heterogeneous Face Recognition Problems

Dictionary.com defines *heterogeneous* as "diverse in kind or nature." In the context of biometrics, heterogeneous face recognition (HFR) is to match face images coming from different modalities.[1] The motivation of the HFR is that face images of the same subject can often be captured by different sensors under different imaging conditions, because of the sensor technology development and broader application requirements.

"In the context of biometrics, heterogeneous face recognition (HFR) is to match face images coming from different modalities."

For example, the sensors can use different spectral bands: visible light spectrum (VIS), near infrared (NIR), and thermal infrared (IR); different content can be acquired: regular two-dimensional (2D) light reflection and three-dimensional (3D) depth data, especially the recently developed RGB-D sensors. Further, the cameras can have different qualities with different prices, for example, high-quality professional cameras, low-quality surveillance or web cameras, or photo scanners; and can be used in different acquisition environments: indoor/outdoor or different weather conditions (sunny, rainy, or snowy).

Therefore, in real applications, the probe and gallery face images may come from different image modalities. For instance, the still face images are usually used for face identity enrollment, while the face images from surveillance video cameras might be used for face matching or search over the still image database.

"Therefore, in real applications, the probe and gallery face images may come from different image modalities."

In this section, various HFR problems are discussed and presented, including both the basic problems that are clearly defined and have been studied in quite a few research works and some other HFR problems that have not been studied extensively.

Basic HFR Problems

The basic heterogeneous face matching problems include VIS vs. Sketch, VIS vs. NIR, VIS vs. 3D, and VIS vs. IR. These specific problems have been clearly defined in previous research works[1][2], and are commonly admitted by researchers in biometrics.

Among the basic and typical heterogeneous face matching problems, VIS vs. Sketch and VIS vs. NIR are the mostly studied in the literature.

There are also approaches performing heterogeneous matching between thermal IR and VIS face images, for example, Li et al.[3], Choi et al.[4], Klare and Jain[2], and approaches to perform recognition between forensic sketches and visible face images[5][2], which is much more challenging than viewed sketches (drawn while viewing), because the drawn sketches can be obtained based on very limited information about the true identity, resulting in the sketches not being similar to the exact person. Compared to the popular VIS vs. Sketch and VIS vs. NIR, there are far fewer publications on VIS vs. 3D and VIS vs. IR matching, although these problems are also defined clearly as heterogeneous face recognition problems.

"Compared to the popular VIS vs. Sketch and VIS vs. NIR, there are far fewer publications on VIS vs. 3D and VIS vs. IR matching, although these problems are also defined clearly as heterogeneous face recognition problems."

There are several reasons why the problems of VIS vs. Sketch and VIS vs. NIR are more popular than others. One is that the high quality 3D range sensors and thermal IR cameras are still expensive, while the acquisition of NIR face images and face sketches does not need to involve expensive sensors. Thus it is relatively easier to collect data for research and practical applications, involving the Sketch, VIS, and NIR images. Another reason could be that it is more challenging to perform VIS vs. 3D or VIS vs. IR, since the image appearance differences between VIS and 3D or VIS and IR are significantly larger than between VIS and Sketch or VIS and NIR. As demonstrated by Goswami et al.[6], some photometric preprocessing of the images can help a lot to get high accuracies for heterogeneous face matching between VIS and NIR modalities. The matching between VIS and Sketch can also have very high accuracies.[7]

The forensic sketches are more challenging than the viewed sketch[5][2]; that is because the forensic sketches drawn by the forensic artists may not know (or the witness may not remember) the "full" face correctly, and thus the limited information can result in the drawn sketches not characterizing the true person well. In other words, it does not really mean that the sketches and VIS are very different modalities.

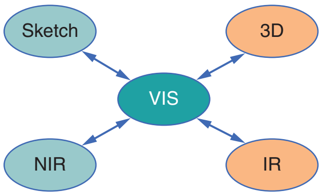

Figure 1: Some typical pairwise, heterogeneous face matching problems (Source: West Virginia University, 2014)

In addition to matching between VIS and other modalities, as shown in Figure 1, there is also heterogeneous matching between any pair of modalities in practice, such as NIR vs. 3D or NIR vs. IR, when the diverse sensors are used more and more in practical applications. To keep the graphic illustration clean, those pairwise matching relations are not shown in Figure 1.

In early studies, researchers usually only dealt with one specific HFR problem, for example, VIS vs. Sketch, while in recent studies, multiple HFR problems were studied to validate the developed methods in different cases.

Not only the basic HFR problems but also some other newly proposed problems can be classified as heterogeneous face matching tasks, which will be introduced next.

Other Heterogeneous Face Matching Problems

Some other face recognition problems in recent studies can be considered as heterogeneous face matching too. These atypical HFR problems include:

1. Matching between face images of different resolutions, that is, high-resolution and low-resolution.[8][9] For this kind of study, some existing face databases were used to "generate" face images at different resolutions. For example, the face images are cropped[9] as 32 32 and then down-sampled to 16 16, 8 8, 6 6, and 4 4. These down-sampled low-resolution face images were up-sampled into 32 32 to mimic the low-resolution face images for their study.
2. Digital photo vs. video frame.[9] Face images can be captured by digital still cameras or extracted from the video sequences captured by video camcorders. The faces from digital photos and video frames can have different resolutions and qualities. Thus face matching between digital photos and video frames can also be considered as a heterogeneous face matching problem.[9]
3. Face recognition with cosmetic changes.[10][11] This can be considered as another heterogeneous face recognition problem. As shown in Figure 2, face images of the same subject may look very different based on whether

"Thus face matching between digital photos and video frames can also be considered as a heterogeneous face matching problem."

Figure 2: Faces with makeup applied (left column) and faces with no makeup (right column) for the same individuals (each row).
(Source: Originally shown in Wen and Guo[11], 2013)

"The matching between face images with or without makeup can be considered as another heterogeneous face recognition problem."

facial makeup is applied or not. The matching between face images with or without makeup can be considered as another heterogeneous face recognition problem.

Actually, it has been found that facial cosmetics can change the perception of faces significantly[12] and can bring a great challenge for face matching computationally.[13][14] Motivated by these studies, we have studied how to address the influence of makeup on face recognition based a dual-attributes approach[11], and a correlation-based approach.[10]

Heterogeneous Face Recognition Algorithms

"The key issue for heterogeneous face matching is how to reduce the difference between heterogeneous face images."

The key issue for heterogeneous face matching is how to reduce the difference between heterogeneous face images. Typically, there exist significant facial appearance changes between heterogeneous face images, even though the face images can be aligned well. The differences can be caused by the variety of sensors (for example, different spectral responses), different image acquisition conditions (for example, by physical devices or hand-drawing), or changes by the subjects themselves (for example, applying facial makeup). So the algorithm development for HFR usually focuses on various approaches to reduce the differences between heterogeneous face images of the same subjects.

Despite the significant progress that has been made for face recognition, most face recognition systems are not designed to handle HFR scenarios currently, including commercial off-the-shelf (COTS) systems. Therefore, there is a need and substantial interest for studying heterogeneous face matching problems.[2]

In this section, some representative approaches to HFR will be presented, based on a grouping into different categories.

Transforming One Modality to Another

To reduce the facial appearance differences between two modalities, one category of approaches is to transform the face images from modality A to another denoted by B, such that face matching can be executed using the "same" modality B approximately. This transformation can be in the raw image level or feature level. If it is in the image level, a new image will be synthesized in modality B, and then the image comparison is likely to use the same modality B; If it is in the feature level, the extracted features from image modality A will be transformed into features in domain B, and then compared to the features extracted directly from image modality B. This kind of approach is typically used to deal with VIS and sketch matching, where a face sketch can be synthesized from a photograph (or vice versa).[15][16][7] There are also some other methods proposed purely for sketch synthesis[17][18], which may be useful for matching VIS and sketch images.

A representative method to sketch synthesis from face photos is the eigen-transform method[15], which is similar to the eigenfaces method[19], but applied to two image modalities. The key idea is the sketch to be synthesized can be reconstructed based on the linear combination of a set of eigenvectors learned from training sketch images, and the combination coefficients are equal to those learned from the corresponding face photo reconstruction. Thus, given a face photo, the reconstruction coefficients can be learned first and then applied to the sketch synthesis from sketch eigenvectors. After synthesis, the pseudo-sketch can be used to match against real sketches in the gallery for recognition.

Other approaches[20][21][22] use the idea similar to image analogies[23] to transform one modality to another, such as NIR to VIS or vice versa. One representative method[20] is to use local patches to build a dictionary for VIS and NIR faces separately and learn a linear combination of the nearest neighbors (similar patches) to reconstruct each patch for a given NIR face image. Then the learned linear reconstruction is applied to a new modality to synthesize a virtual VIS face for matching with other VIS images in the gallery.

Photometric Preprocessing

The second category of approaches to HFR is to use photometric preprocessing techniques to normalize the lighting or illumination in face images of each modality so that the differences between heterogeneous face images can be reduced. These preprocessing methods were originally developed to deal with illumination changes in visible light face images, but were then adapted to address the heterogeneous face matching problems, such as VIS vs. NIR face images. For these approaches, the underlying assumption is that the heterogeneity of face images is caused by the lighting or reflection differences in face surfaces.

"The second category of approaches to HFR is to use photometric preprocessing techniques to normalize the lighting or illumination in face images of each modality so that the differences between heterogeneous face images can be reduced."

Goswami et al.[6] gave a good summary of different photometric preprocessing techniques for HFR. Typically there are three different methods for photometric preprocessing, which will be introduced here:

One method is called sequential chain (SQ) preprocessing. It uses a series of steps for face image preprocessing. First, the Gamma correction is executed, which enhances the local dynamic range of the face image in darker regions, while compressing the range in bright and highlight regions. Second, the Difference of Gaussian (DoG) filtering is performed to compress the low frequency or nonessential information while maintaining or enhancing the gradient information that is more useful for recognition. Third, contrast equalization is used to rescale the intensity values globally and reduce the possibility of having extreme values during the processing in previous steps.

Another method is called single scale retinex (SSR). Usually the image intensity value, I, can be modeled as the product of illumination L and surface reflectance R. In the SSR method, the illumination component L is estimated by using the blurred image computed from the original face image. For example, the Gaussian filter can be used to compute the blurred image. Then the reflectance component R can be estimated by subtracting the illumination component from the original image in the logarithm domain. The SSR is applied to different modality images separately to compute the reflectance. The resulted reflectance images are assumed to be similar for heterogeneous face images, and are then used for feature extraction and matching.

The third method is called self quotient image (SQI). The SQI is very similar to the SSR operation. It is defined by the ratio between the original face image and a smoothed version of the original image, without using the logarithm computation. The ratio image is then used for feature computation and matching, replacing the original face image.

"Currently the photometric preprocessing methods are mainly used for VIS vs. NIR face images."

Currently the photometric preprocessing methods are mainly used for VIS vs. NIR face images. As shown in Figure 3, various photometric preprocessing methods can make the NIR and VIS face images look more similar. However, it is not clear if these methods are useful or not for other heterogeneous face matching problems, such as VIS vs. IR or VIS vs. 3D.

Figure 3: The effect of photometric preprocessing on heterogeneous face images (top: VIS, bottom: NIR); left to right: raw images, SQ, SQI, and SSR processing results.
(Source: Originally shown in Goswami et al.[6])

Another issue is that even though the photometric preprocessing can make the face images similar, it still needs feature mapping or other learning methods to further improve the performance for HFR in practice.

Common Subspace Projection

The third category of HFR approaches is to generate common subspaces so that both modalities of face images can be projected into, and the differences between heterogeneous images are expected to be minimized after the projection, as illustrated in Figure 4. New features can be generated after the joint projections into the common space.

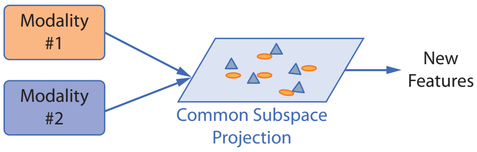

Figure 4: The common subspace projection to build the relationship between two different modalities of data and generate new features to minimize the differences
(Source: West Virginia University, 2014)

Classical methods to generate the common subspaces include the canonical correlation analysis (CCA)[24], and partial least squares (PLS).[25] These methods and their kernel versions for nonlinear mapping have been used for HFR, for example, by Sharma and Jacobs[8], Yang et al.[26], and Yi et al.[27]

Given face images corresponding to two different modalities, the CCA method can learn a pair of directions to maximize the correlation of the original data in the new subspace. The PLS is to learn a latent subspace such that the covariance between latent scores of the data from two modalities is maximized. Both the CCA and PLS methods can have linear mapping and kernel based extensions for nonlinear mapping.

"Both the CCA and PLS methods can have linear mapping and kernel based extensions for nonlinear mapping."

In addition to the classical methods, there are some other recent approaches to compute the common subspace in different ways. For example, Lin and Tang[28] proposed a method called Common Discriminant Feature Extraction (CDFE) for inter-modality face recognition. Two transforms are simultaneously learned to transform the samples in both modalities respectively to the common feature space. The learning objective incorporates both the discriminative power and local smoothness of the feature transformation.

Another method is the coupled discriminant analysis (CDA) by Lei et al.[9], which incorporates constraints such as locality information of the features and discriminative computation similar to the classical linear discriminant analysis (LDA), to improve the performance for heterogeneous face matching. More recently, the kernel-prototype–based similarity measure for HFR[2] was proposed, which pursues the kernel trick by Balcan et al.[29] to represent each face image with a set of training images, serving as prototypes.

Random Subspaces

The random subspace (RS) method by Ho[30] was developed to deal with the small sample size problem in recognition, using the idea similar to the classical bagging[31] and random forests[32] methods. The RS method is also useful to improve and generalize the classification performance, based on sampling a subset of features and classifier training in the reduced feature space. Then multiple classifiers can be learned from the multiple sets of randomly sampled features. These classifiers can be combined together to form a much stronger classifier or recognizer.

"The random subspace (RS) method by Ho was developed to deal with the small sample size problem in recognition, using the idea similar to the classical bagging and random forests methods."

Wang and Tang[33] used the random subspace with linear discriminant analysis (LDA) called RS-LDA for visible light face recognition. Klare and Jain[34] adapted RS-LDA for heterogeneous face recognition, by using multiple samplings of face patches from both VIS and NIR face images. The random subspace is also extended to the kernel prototype similarity measures[2] for HFR.

Dual Attributes

Attributes are a semantic level description of visual traits, as discussed, for instance, by Lampert et al.[35] and Farhadi et al.[36] For example, a horse can be described as four legged, mammal, can run, can jump, and so on. A nice property of using attributes for object recognition is that the basic attributes might be learned from other objects, and shared among different categories of objects.[37]

Facial attributes are a semantic level description of visual traits in faces, such as big eyes, or a pointed chin. Kumar et al.[38] showed that a robust face verification can be achieved using facial attributes, even if the face images are collected from uncontrolled environments over the Internet.

Motivated by the usefulness of facial attributes, a method called dual attributes was recently proposed by Wen and Guo[11] for face verification robust to facial appearance changes caused by the makeup. The key idea is that the dual attributes can be learned from faces with and without cosmetics, separately. Then the shared attributes can be used to measure facial similarity irrespective of cosmetic changes. In essence, dual attributes are capable of matching faces with or without makeup in a semantic level, rather than a direct matching with low-level features.

"The key idea is that the dual attributes can be learned from faces with and without cosmetics, separately."

The dual attributes method by Wen and Guo[11] may be adapted to other heterogeneous face matching problems.

Multiview Discriminative Learning

In the methods introduced above, typically only two modalities are used for HFR. Is it possible to deal with multiple modalities in the formulation? The answer is yes.

For example, the CCA can be extended to a multiview CCA by Rupnik and Shawe-Taylor.[39] Another way is to use the principle of LDA to derive a so-called multiview discriminant analysis (MDA) method by Kan et al.[40] It learns multiple view-specific linear transforms in a non-pairwise manner by optimizing a generalized Rayleigh quotient, that is, maximizing the between-class variations and minimizing within-class variations in a low dimensional subspace. The optimization problem is then solved by using the generalized eigenvalue decomposition technique.

Another method is the generalized multiview analysis by Sharma et al.[41], where the cross-view correlation is obtained from training examples corresponding to the same subjects or identities. This correspondence requirement is not needed in the MDA formulation.[40]

These multiview analysis methods[40][41] have been shown to be useful for some heterogeneous image matching problems, such as photo vs. sketch and VIS vs. NIR.

"These multiview analysis methods have been shown to be useful for some heterogeneous image matching problems, such as photo vs. sketch and VIS vs. NIR."

Heterogeneous Face Databases

To facilitate the study of heterogeneous face recognition, several databases have been assembled. A summary of the existing databases are presented in this section.

CUFS Database (Sketch-VIS)

This database was collected by the Chinese University of Hong Kong. The CUHK Face Sketch Database contains 606 subjects with VIS and sketch face pairs.[7] There are 1,216 images in total. This is probably the first publicly available database for heterogeneous face matching.

CUFSF Database (Sketch-VIS)

This is an extended version of the CUFS database, containing 1,194 subjects with 2,388 image pairs of VIS and sketch by Zhang et al.[42] The sketch photos were drawn by artists when viewing the original face images for each subject. It is called *viewed sketches* by Klare and Jain[2] in contrast to the forensic sketches.

CASIA-HFB Database (VIS-NIR-3D)

This is probably the first database that contains more than two face modalities, assembled from the Institute of Automation, Chinese Academy of Sciences (CASIA) by Li et al.[43] It has 100 subjects of 992 face images in total. Each subject has four VIS, four NIR, and one or two 3D face images. The cropped face images were provided with the eye coordinates aligned manually. Some baseline results were provided based on direct matching with the classical PCA and LDA features. Later on, the database was extended to 202 subjects just for the VIS and NIR image modalities, resulting in 5,097 face images for VIS and NIR modalities.

"This is probably the first database that contains more than two face modalities, assembled from the Institute of Automation, Chinese Academy of Sciences…"

Cross-Spectral Dataset (VIS-NIR)

This dataset by Goswami et al.[6] contains VIS and NIR face pairs for 430 subjects over multiple sessions, collected from the University of Surrey in the United Kingdom. Different pose angles in pitch and yaw directions were captured for every 10 degrees. Each subject has at least three poses. In total, there are 2,103 NIR and 2,086 VIS face images. Twelve algorithms were provided as the baseline results together with the database, based on the combination of different photometric preprocessing methods, features, and matching techniques.

LDHF-DB (VIS-NIR, Long Distance)

This database by Maeng et al.[44] was collected by the Korea University. It contains 100 subjects at different distances to the cameras. Each subject was captured at distances of 60, 100, and 150 meters, separately, using both VIS and NIR cameras. There are 1,600 face images in total. This dataset emphasizes the long distance acquisition of heterogeneous face images.

"This dataset emphasizes the long distance acquisition of heterogeneous face images."

UND Database (VIS-IR)

The database contains 82 subjects with multiple IR and VIS face images for each subject. The total number of face images in this database is 2,292. It was used by Choi et al.[4] for IR to VIS face recognition.

NPU3D Database (VIS-3D)

The NPU3D database by Zhang et al.[45] contains Chinese VIS and 3D faces, collected at Northwestern Polytechnical University, China, using the Konica Minolta Vivid 910 3D laser scanner. The acquisition distance is about 1.5 meters. There are 300 individuals captured with 35 different scans (various pose, facial expression, accessory and occlusion) per subject. In total, there are 10,500 3D facial surface scans with the corresponding VIS images.

CASIA NIR-VIS 2.0 Database (VIS-NIR)

It contains 725 subjects of 17,580 face images from multiple recording sessions, in which the first session is identical to the CASIA-HFB database. Each subject has 1–22 VIS and 5–50 NIR face images. Different evaluation protocols were also provided with the database as well by Li et al.[46]

Other Databases

There are also some other databases that are either small, seldom used, or just private, such as, for example, the VIS and IR face database collected by the Pinellas County Sheriff's Office and forensic sketches and VIS databases, introduced by Klare and Jain[2].

Some Thoughts on Future Directions

"As an emerging topic in biometrics, HFR has attracted more and more attention recently."

As an emerging topic in biometrics, HFR has attracted more and more attention recently. However, the study of HFR is still in its early stage, and more efforts are needed to advance the field of research. Here some new thoughts are presented, hopefully to inspire new efforts to address the challenging research on HFR.

Identify Which Methods Can Work on Which HFR Problems

There are different modalities to match within HFR, such as Sketch vs. VIS, NIR vs. VIS, and so on. Different algorithms and approaches have been developed, which are typically for one specific HFR problem or two, but not for all. Even though an algorithm can be tested on different HFR problems experimentally, the recognition accuracies could be very different for different HFR problems. For example, an algorithm can get 95-percent accuracy on VIS vs. sketch, but may only achieve 60-percent accuracy when applied to VIS vs. IR. So an issue is raised: which methods can work on which HFR problems? New investigations can be performed to address this issue, and then one can know which methods are appropriate to solve what kinds of HFR problems. It is especially important for real applications of biometrics systems, not just for academic research. A systematic evaluation of the existing (and future) algorithms on each of the HFR problems could be done towards addressing this issue.

Deal with the Degrees of Heterogeneity in HFR

Related to the previous issue, another is to study and define the degrees of heterogeneity in various heterogeneous face matching problems. As presented earlier, there are a variety of HFR problems. However, it has not yet been studied just how heterogeneous it could be between two given modalities of face images. By defining and measuring the degrees of heterogeneity, one can know just how difficult it is to solve a specific HFR problem: the more heterogeneous, the more difficult to address typically.

"By defining and measuring the degrees of heterogeneity, one can know just how difficult it is to solve a specific HFR problem: the more heterogeneous, the more difficult to address typically."

Further, when a new HFR problem is proposed, one can predict how difficult it will be to address it before developing an algorithm to solve it, based on the measure of degrees of heterogeneity. The challenge is how to define and measure the degree of heterogeneity universally over different matching problems.

And also, the measure of the heterogeneity can help classify the existing (and future) algorithms into different categories based on their capabilities to address the HFR problems at different levels of heterogeneity.

Explore New Learning Methods to Solve HFR Problems

As stated above, the study of HFR is still not mature; new algorithms are expected to be developed to improve the recognition performance. In developing new algorithms, one promising direction is to explore learning-based methods. Since it is difficult (if not impossible) to model how the image appearance is changed from one modality to another, example-based learning approaches are probably the only way to study the differences between two modalities and to build the relations between face images in two modalities.

"In developing new algorithms, one promising direction is to explore learning-based methods."

In exploring learning-based methods, one direction is to study the domain adaptation methods to adapt the data from one modality to another. Recently, we have shown that the adaptive support vector machines (A-SVM) by Yang et al.[47] can be applied for action recognition from VIS to IR by Zhu and Guo[48]. Based on this, we can expect that the A-SVM or other domain adaptation methods could be helpful to address the HFR problems.

Collect Larger Databases with Public Access

As stated earlier, some HFR databases have been assembled; however, few of them are large, compared to the homogeneous (same modality) face recognition databases. By collecting larger databases, one can evaluate the algorithm's performance better towards real applications. Further, there are fewer databases for VIS vs. 3D, VIS vs. IR, makeup vs. no makeup, or containing multiple modalities for the same subjects. New databases can be collected to facilitate the study of various HFR problems, rather than just VIS vs. Sketch or VIS vs. NIR.

"By collecting larger databases, one can evaluate the algorithm's performance better towards real applications."

Other HFR Problems

Some typical and atypical HFR problems were introduced earlier. However, new HFR problems can still be identified along with new sensor development or acquisition environment changes.

Further, some existing face recognition problems may be revisited by considering them as HFR. In this way, new ideas may be inspired to address the well-defined problems from new angles. For example, human aging can cause significant facial appearance changes, as shown in Figure 5. Cross-age face recognition is a well-defined, challenging problem. Various methods have been proposed, such as the generative approaches based on age synthesis by Gong et al.[49], Wu and Chellappa[50], Park et al.[51], Ramanathan and Chellappa[52], and discriminative approaches by Yadav et al.[53], Li et al. [54], Ling et al.[55], and Biswas et al.[56] Because of the space limit, it will not be discussed in detail here, but the cross-age face recognition can be considered as a HFR problem as well.

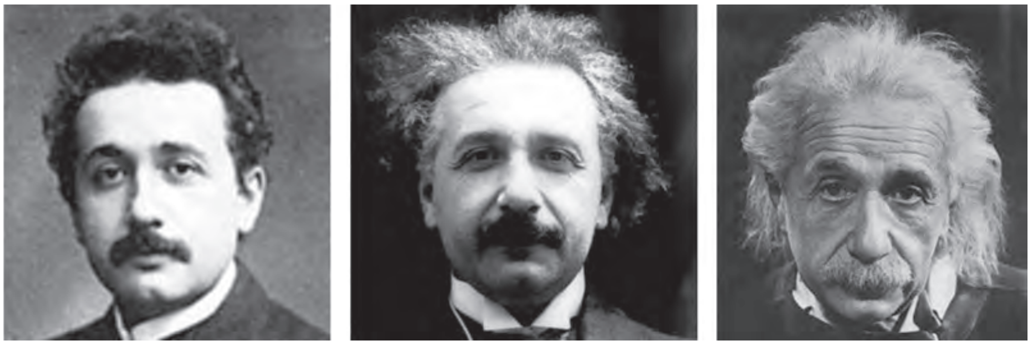

Figure 5: Aging can cause significant facial appearance changes.
(Source: Image Search over the Internet, 2014)

Concluding Remarks

An emerging topic in biometrics, called heterogeneous face recognition, has been presented. Several specific HFR problems, both typical and atypical, have been introduced. Some representative approaches to HFR have been described based on a categorization. Various HFR databases have been listed to researchers, and some new thoughts on future exploration of HFR have been introduced as well. Hopefully this article will inspire new research efforts to address the challenging and interesting heterogeneous face recognition problems.

"Hopefully this article will inspire new research efforts to address the challenging and interesting heterogeneous face recognition problems."

Complete References

[1] Li, S., *Encyclopedia of Biometrics* (Springer: 2009).

[2] Klare, B. and A. Jain, "Heterogeneous Face Recognition Using Kernel Prototype Similarities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, No. 6, pp. 1410–1422, 2013.

[3] Li, J., P. Hao, C. Zhang, and M. Dou, "Hallucinating Faces from Thermal Infrared Images," *Proc. Int'l Conf. Image Processing*, pp. 465–468, 2008.

[4] Choi, J., S. Hu, S. Young, and L. Davis. "Thermal to Visible Face Recognition." *Proc. of SPIE*, Vol. 8371, pages 83711L–1, 2012.

[5] Klare, B., Z. Li, and A. Jain, "Matching Forensic Sketches to Mugshot Photos," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 639–646, Mar. 2011.

[6] Goswami, D., C. Chan, D. Windridge, and J. Kittler, "Evaluation of face recognition system in heterogeneous environments (Visible vs NIR)," *2011 IEEE International Conference on Computer Vision Workshops*, pages 2160–2167.

[7] Wang, X. and X. Tang, "Face Photo-Sketch Synthesis and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, Nov. 2009.

[8] Sharma, A. and D. Jacobs. "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch," *CVPR*, pages 593–600, 2011.

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

[10] Guo, G-D., L. Wen, and S. Yan, "Face authentication with makeup changes," *IEEE Trans. Circuits and Systems for Video Technology*, DOI:10.1109/TCSVT.2013.2280076

[11] Wen, L. and G-D. Guo, "Dual attributes for face verification robust to facial cosmetics," *Journal of Computer Vision and Image Processing*, NWPJ-201301-82, Vol. 3, No. 1, pages 63–73, 2013.

[12] Ueda, S. and T. Koyama, "Influence of make-up on facial recognition," *Perception*, 39(2):260, 2010.

[13] Dantcheva, A., C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?" *IEEE Conf. on Biometrics: Theory, Applications and Systems*, Washington DC, USA, 2012.

[14] Chen, C., A. Dantcheva, and A. Ross, "Automatic facial makeup detection with application in face recognition," *Proc. of International Conference on Biometrics* (ICB), Madrid, Spain, June 2013.

[15] Tang, X. and X. Wang, "Face Sketch Recognition," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 50–57, Jan. 2004.

[16] Liu, Q., X. Tang, H. Jin, H. Lu, and S. Ma, "A Nonlinear Approach for Face Sketch Synthesis and Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1005–1010, 2005.

[17] Gao, X., J. Zhong, J. Li, and C. Tian, "Face Sketch Synthesis Algorithm Based on E-HMM and Selective Ensemble," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 487–496, Apr. 2008.

[18] Zhang, W., X. Wang, and X. Tang, "Lighting and Pose Robust Face Sketch Synthesis," *Proc. European Conf. Computer Vision*, 2010.

[19] Turk, M. and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[20] Chen, J., D. Yi, J. Yang, G. Zhao, S. Z. Li, and M. Pietikainen, "Learning mappings for face synthesis from near infrared to visual light images," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 156–163, 2009.

[21] Wang, R., J. Yang, D. Yi, and S. Z. Li, "An analysis-by-synthesis method for heterogeneous face biometrics," *Advances in Biometrics* (Berlin: Springer 2009), pp. 319–326.

[22] Liu, M., W. Xie, X. Chen, Y. Ma, Y. Guo, J. Meng, and Q. Qin, "Heterogeneous face biometrics based on Gaussian weights and invariant features synthesis," *IEEE 2nd International Conference on Computing, Control and Industrial Engineering (CCIE)*, Vol. 2, pp. 374–377, 2011.

[23] Hertzmann, A., C. Jacobs, N. Oliver, B. Curless, and D. Salesin, "Image analogies," *SIGGRAPH*, 2001.

[24] Hotelling, H., "Relations between two sets of variates." *Biometrika* 28, 321–377 (1936).

[25] Wold, H. "Partial least squares," in *Encyclopedia of Statistical Sciences,* edited by S. Kotz and N. Johnson, volume 6, pages 581–591. Wiley, New York, 1985.

[26] Yang, W., D. Yi, Z. Lei, J. Sang, and S. Li, "2D-3D face matching using CCA," *IEEE International Conference on Automatic Face Gesture Recognition*, pages 1–6, 2008.

[27] Yi, D., R. Liu, R. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," *IAPR International Conf. on Biometric*, pages 523–530, 2007.

[28] Lin, D. and X. Tang, "Inter-Modality Face Recognition," *Proc. European Conf. Computer Vision*, pages 13–26, 2006.

[29] Balcan, M.-F., A. Blum, and S. Vempala, "Kernels as Features: On Kernels, Margins, and Low-Dimensional Mappings," *Machine Learning*, vol. 65, pp. 79–94, 2006.

[30] Ho, T. K., "The Random Subspace Method for Constructing Decision Forests," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832–844, 1998.

[31] Breiman, L., "Bagging Predictors," *Machine Learning*, vol. 24, pp. 123–140, 1996.

[32] Breiman, L., "Random forests," *Machine Learning* 45, No. 1 (2001): 5–32.

[33] Wang, X. and X. Tang, "Random Sampling for Subspace Face Recognition," *Int'l J. Computer Vision*, vol. 70, no. 1, pp. 91–104, 2006.

[34] Klare, B. and A. Jain, "Heterogeneous face recognition: Matching NIR to visible light images," *Int'l Conf. on Pattern Recognition*, 2010, pp. 1513–1516.

[35] Lampert, C., H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 951–958, 2009.

[36] Farhadi, A., I. Endres, D. Hoiem, and D. Forsyth, "Describing objects by their attributes," *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1778–1785, 2009.

[37] Farhadi, A., I. Endres, and D. Hoiem, "Attribute-centric recognition for cross-category generalization," *IEEE Conf. on CVPR*, pages 2352–2359, 2010.

[38] Kumar, N., A. Berg, P. Belhumeur, and S. Nayar, "Attribute and simile classifiers for face verification," *IEEE International Conf. on Computer Vision*, pages 365–372, 2009.

[39] Rupnik, J. and J. Shawe-Taylor, "Multi-view canonical correlation analysis," *SiKDD*, 2010.

[40] Kan, M., S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *European Conf. on Computer Vision*, pp. 808–821 (2012).

[41] Sharma, A., A. Kumar, H. Daume III, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," IEEE Conference on Computer Vision and Pattern Recognition (2012).

[42] Zhang, W., X. Wang, and X. Tang, "Coupled Information-Theoretic Encoding for Face Photo-Sketch Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2011.

[43] Li, S. Z., Z. Lei, and M. Ao, "The HFB face database for heterogeneous face biometrics research." *IEEE Computer Vision and Pattern Recognition Workshops*, pp. 1–8, 2009.

[44] Maeng, H., S. Liao, S.-W. Lee, and A. K. Jain. "Nighttime face recognition at long distance: cross-distance and cross-spectral matching," *Asian Conf. on Computer Vision*, pp. 708–721. 2012.

[45] Zhang, Y., Z. Guo, Z. Lin, H. Zhang, and C. Zhang, "The NPU Multi-case Chinese 3D Face Database and Information Processing," *Chinese Journal of Electronics*, vol. 21, no. 2 (2012): 283–286.

[46] Li, S. Z., D. Yi, Z. Lei, and S. Liao, "The CASIA NIR-VIS 2.0 face database," *Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 348–353, 2013.

[47] Yang, J., R. Yan, and A. Hauptmann, "Cross-Domain Video Concept Detection using Adaptive SVMs," *Proc. MM*, 2007.

[48] Zhu, Y. and G-D. Guo, "A study on visible to infrared action recognition," *IEEE Signal Processing Letters*, Vol. 20, No. 9, pages 897–900, 2013.

[49] Gong, D., Z. Li, D. Lin, J. Liu, and X. Tang, X. "Hidden Factor Analysis for Age Invariant Face Recognition," *ICCV* 2013.

[50] Wu, T. and R. Chellappa, "Age invariant face verification with relative craniofacial growth model," *Computer Vision–ECCV 2012* (pp. 58–71).

[51] Park, U., Y. Tong, and A. K. Jain, "Age-invariant face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(5), 947–954.

[52] Ramanathan, N. and R. Chellappa, "Face verification across age progression," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3349–3361, 2006.

[53] Yadav, D., M. Vatsa, R. Singh, and M. Tistarelli, "Bacteria Foraging Fusion for Face Recognition across Age Progression," *Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2013) pp. 173–179.

[54] Li, Z., U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE Transactions on Information Forensics and Security*, (2011) 6(3), 1028–1037.

[55] Ling, H., S. Soatto, N. Ramanathan, and D. W. Jacobs, "Face verification across age progression using discriminative methods," *IEEE Transactions on Information Forensics and Security*, (2010) 5(1), 82–91.

[56] Biswas, S., G. Aggarwal, and R. Chellappa, "A non-generative approach for face recognition across aging," in IEEE Second Int'l Conf. on Biometrics: Theory, Application and Systems, 2008.

Author Biography

Guodong Guo received his BE degree in Automation from Tsinghua University, Beijing, China, in 1991, a PhD in Pattern Recognition and Intelligent Control from the Chinese Academy of Sciences, in 1998, and a PhD in computer science from the University of Wisconsin-Madison, in 2006. He is currently an assistant professor in the Lane Department of Computer Science and Electrical Engineering at West Virginia University. In the past, he has visited and worked in several places, including INRIA, Sophia Antipolis, France, Ritsumeikan University, Japan, Microsoft Research, China, and North Carolina Central University. He won the North Carolina State Award for Excellence in Innovation in 2008, and Outstanding Researcher (2013–2014) and New Researcher of the Year (2010–2011) at CEMR, WVU. He was selected as the "People's Hero of the Week" by BSJB under MMTC on July 29, 2013. His research areas include computer vision, machine learning, and multimedia. He is the author of *Face, Expression, and Iris Recognition Using Learning-based Approaches* (2008), co-editor of *Support Vector Machines Applications* (2014), and has published over 60 technical papers in face, iris, expression, and gender recognition, age estimation, and multimedia information retrieval. He can be contacted at Guodong.Guo@mail.wvu.edu