Accepted Manuscript

Super-Resolution for Biometrics: A Comprehensive Survey

Kien Nguyen, Clinton Fookes, Sridha Sridharan, Massimo Tistarelli,
Mark Nixon

PII: S0031-3203(18)30004-9
DOI: 10.1016/j.patcog.2018.01.002
Reference: PR 6411

To appear in: *Pattern Recognition*

Received date: 20 April 2017
Revised date: 18 December 2017
Accepted date: 7 January 2018

Please cite this article as: Kien Nguyen, Clinton Fookes, Sridha Sridharan, Massimo Tistarelli, Mark Nixon, Super-Resolution for Biometrics: A Comprehensive Survey, *Pattern Recognition* (2018), doi: 10.1016/j.patcog.2018.01.002

Highlights

- Discuss the significance of super-resolution in the context of biometrics and how it is different from the general super-resolution.
- Present a background knowledge in general super-resolution approaches.
- Provide a comprehensive review of existing super-resolution approaches for biometric modalities, including face (2D and 3D), iris, gait and latent prints (fingerprint and palmprint) and other emerging modalities.
- Highlight and discuss current challenges and recommendations for future research in Super-resolution for biometrics.

Super-Resolution for Biometrics: A Comprehensive Survey

Kien Nguyen[a,*], Clinton Fookes[a], Sridha Sridharan[a], Massimo Tistarelli[b], Mark Nixon[c]

[a] *Image and Video Lab, Queensland University of Technology, Queensland, Australia*
[b] *Computer Vision Laboratory, University of Sassari, Sassari, Italy*
[c] *School of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom*

Abstract

The lack of resolution of imaging systems has critically adverse impacts on the recognition and performance of biometric systems, especially in the case of long range biometrics and surveillance such as face recognition at a distance, iris recognition and gait recognition. Super-resolution, as one of the core innovations in computer vision, has been an attractive but challenging solution to address this problem in both general imaging systems and biometric systems. However, a fundamental difference exists between conventional super-resolution motivations and those required for biometrics. The former aims to enhance the visual clarity of the scene while the latter, more significantly, aims to improve the recognition accuracy of classifiers by exploiting specific characteristics of the observed biometric traits. This paper comprehensively surveys the state-of-the-art super-resolution approaches proposed for four major biometric modalities: face (2D+3D), iris, fingerprint and gait. We approach the super-resolution problem in biometrics from several different perspectives, including from the spatial and frequency domains, single and multiple input images, learning-based and reconstruction-based approaches. Especially, we highlight two special categories: feature-domain super-resolution which performs super-resolution directly on the feature space to purposely improve the recognition performance, and deep-learning super-resolution which discusses the most recent advances in deep learning for the super-resolution task. Finally, we discuss the current and open research challenges and provide recommendations into the future for the improved use of super-resolution with biometrics.

Keywords: Super-resolution, Biometrics, Face Recognition, Iris Recognition, Gait Recognition, Fingerprint Recognition, Non-ideal biometrics, Human identification at a distance, Deep learning

1. Introduction

In the digital imaging area, image resolution describes the details contained in an image: the higher the resolution, the higher pixel density and the richer the visual detail. In most applications, high-resolution (HR) images are desired and often required because they provide both sharp and clear pictorial information for human perception and rich details for automatic machine interpretation and representation. However, the spatial resolution is limited according to the imaging sensors

*Corresponding author
Email addresses: `k.nguyenthanh@qut.edu.au` (Kien Nguyen), `c.fookes@qut.edu.au` (Clinton Fookes), `s.sridharan@qut.edu.au` (Sridha Sridharan), `tista@uniss.it` (Massimo Tistarelli), `msn@ecs.soton.ac.uk` (Mark Nixon)

and optics employed in the capture (lens blur, lens aberration effects, aperture diffraction and optical blurring due to motion), which generates low-resolution (LR) observation image versions of the scene of interest. LR images are prevalent in real-life applications such as surveillance due to the cheap LR imaging systems available for use in those particular environments and the long distance between the subjects of interest and the camera [1, 2, 3]. The resolution drop decreases the amount of information available for identifying or verifying an individual, ultimately resulting in a severe degradation for both human perception and machine interpretation. To address this, Super-Resolution (SR) techniques have been proposed and employed to overcome the inherent resolution limitation of the current imaging systems and improve the performance of most digital image processing applications.

Super-resolution is a technique reconstructing a high-resolution image from one or multiple low-resolution observation images of the same scene, thereby increasing the high-frequency components and removing the degradation caused by the imaging process of the low-resolution camera [4, 5]. SR has been one of the most spotlighted areas for the research community (searching the keyword "super resolution" returns 95,496 results in the ScienceDirect database and 93,443 results in Springerlink). As a core problem in computer vision and image processing, SR has enjoyed great success in many applications, including: satellite and aerial imaging [6, 7, 8], medical image processing [9, 10, 11], automated mosaicking [12, 13], compressed image/video enhancement [14, 15, 16], action recognition [17, 18], pose estimation [19, 20] and biometrics [21, 22, 23].

Biometrics, defined as a reliable method for automatic identification/verification of individuals based on their physiological and behavioral characteristics [24, 25, 1], has witnessed a tremendous surge in activity over the last decade. Despite the current success, the major challenge of the existing biometric systems is the short distance of image acquisition, e.g. touching a fingerprint sensor, staying still less than 60 cm from iris cameras or several meters from face cameras [1]. There exists a real need to increase and expand the capture volume to broaden the application domain and relax the distance constraints imposed on the participants. This, at the same time, raises new challenges in the lack of resolution and image quality. Employing advanced hardware solutions such as HR imaging sensors and long focal lenses is not always possible, and in these situations the lack of resolution in the biometrics systems can be addressed from a signal processing point of view by SR techniques. SR techniques enhance both the resolution and quality of the acquired images, enabling the existing biometric systems to extend their application domain to less-constrained (unobtrusive) and non-ideal imaging conditions (long distance between the camera and the subject, occlusion). This is critical in applications such as surveillance and suspect detection and monitoring.

Even though there are a number of survey papers on SR [4, 26, 27, 28, 5], none of these focus on the specific requirements and implications on the use of SR in the biometrics domain. While SR can be regarded as a general technique of image processing, the use of it to improve the accuracy in biometrics is not straightforward in many scenarios. Researchers have proposed innovations in the application of super-resolution that are specific to each biometric modality such as face, iris, gait and fingerprint to improve not only the visual perception but also the recognition performance of the biometric systems. This paper provides a comprehensive survey of SR techniques in biometrics, addressing major categories, major developments and state-of-the-art approaches. Face, iris, gait and fingerprint are the main modalities to be investigated due to their widespread adoption.

The rest of this paper is organized as follows: Section 2 introduces the fundamental concepts of super-resolution, Section 3 surveys state-of-the-art SR approaches proposed for face (2D and 3D), iris, gait and other modalities. Section 4 discusses the current states, open issues and trends in SR for biometrics. The paper is concluded in Section 5.

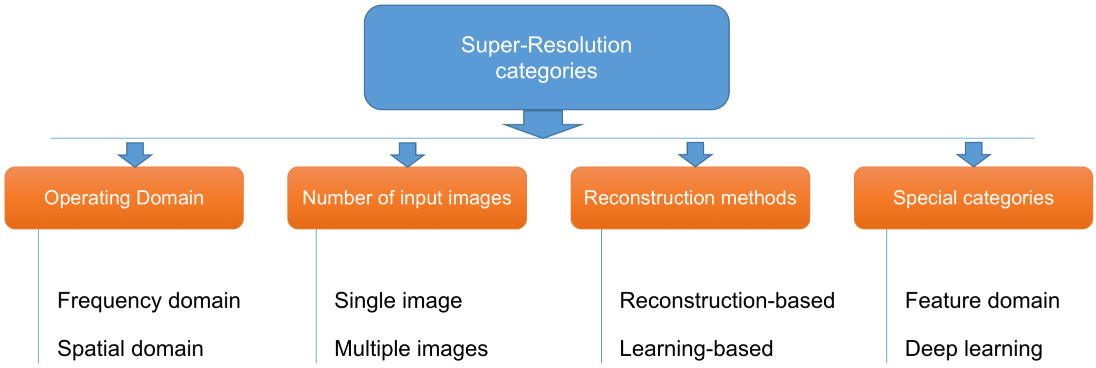


Figure 1: Super-resolution approaches can be categorized based on different factors: operating domain, number of input images, reconstruction methods and special categories for recognition improvement.

2. Super-resolution

The pioneering work in Super-Resolution dated back to 1974 by Gerchberg [29], where he showed that resolution of a data object could be greatly improved through error energy reduction. The author proposed an iterative phase retrieval approach to recover the high frequencies from a single LR image. Since then, the SR technique has been investigated and widely applied in a wide range of applications. In general, SR is an inherently ill-posed problem since the details presented in the low-resolution observations are usually insufficient to reconstruct the original high-resolution image uniquely. Solving the super-resolution problem consists of three main tasks: an alias-free up-sampling of the image, thereby increasing the maximum spatial frequency; removing degradations that arise during the image capturing, e.g. blur and noise; and registration and fusion of multiple samples [30]. The methods in SR differ widely due to the choice of domain, the choice of the number of LR images required and the choice of reconstruction methods. SR approaches in the literature can be categorized based on these factors as follows.

- Operating domain: Spatial domain vs. Frequency domain
  - Spatial domain: these approaches perform super-resolution on the pixel intensity values, which directly corresponds to a visual related enhancement. Numerous spatial domain reconstruction-based methods have also been proposed [31, 32, 33, 34, 35, 36]. Fundamentally, these methods try to model a wide range of motions and degradations and include prior knowledge for regularization. The flexibility, however, comes at the cost of increased computational complexity.
  - Frequency domain: these approaches capitalize on the aliasing that exists in the LR images, an effect easily modeled in the frequency domain. The frequency domain super-resolution algorithms are superior to spatial domain methods in their theoretical simplicity, resulting in most early attempts in SR being processed in the frequency domain [29, 37, 38, 39]. These methods also have low computational complexity and are suitable for parallel implementation due to the simple decoupling of the frequency domain equations. However, the principal limitation of these techniques is that they are limited to using a global translation in the observation model.

- Number of input images: Multiple images vs. Single image
  - Multiple images: a major condition for efficient reconstruction from multiple LR images is that they contain complementary information, which may be caused by such effects as sub-pixel shift and different viewpoints. The high frequencies in the resulting image, which represent the newly-learned details, are in fact available in the measurements in an aliased form. The SR process recovers these high frequencies by exploiting multiple LR images, each exhibiting a different aliasing effect. This explains why such resolution improvement is possible in the first place. However, to enable the recovery of the aliased frequencies uniquely, a sufficient number of low-resolution images are required.
  - Single image: while multiple-image super-resolution techniques capitalize on the complementary information in multiple inputs to reconstruct the high-resolution image, single-image super-resolution techniques only use one single input for this task. The extra information is sought from prior knowledge, such as imaging models and training data. Hence for both multiple- or single-image SR, the extra information source may come from other LR images of the same scene or from prior knowledge.
- Reconstruction methods: Reconstruction-based vs. Learning-based
  - Reconstruction-based: these approaches attempt to *recover* the lost high frequency components. The reconstruction-based approaches seek the lost high frequency components in multiple complementary LR images [40, 41, 42, 43]. When rich and complementary details exist, the reconstruction-based approaches are extremely accurate.
  - Learning-based: these approaches attempt to *learn* the lost high frequency components instead. The learning-based approaches seek the lost high frequency components from the training datasets [44, 45, 46, 47, 48]. Applying this to biometrics is challenging since the learned information, in some cases, may introduce spurious high frequencies, which may improve the human perception but is detrimental to the recognition accuracy.
- Special categories:
  - Feature-domain or recognition-oriented domain: super-resolution in these approaches is performed on the feature domain to purposely improve the recognition performance rather than the visual appearance. These approaches shift the super-resolution operation from the pixel domain to the feature domain which is used directly for the recognition stage [49, 50, 51, 23, 52].
  - Deep Learning based: recent advances in deep learning theory allow automatic and end-to-end learning of the LR-HR mapping from the training data. This mapping can be represented as a deep neural network, when training end-to-end, outperforming hand-crafted models in other approaches [53, 54, 55, 56, 57, 58].

A summary of the aforementioned categories is presented in Figure 1. Even though there are numerous strategies to implement SR techniques, a fundamental SR approach can be summarized in 3 tasks: designing observation model, registering/aligning LR images and reconstructing the original HR image. These tasks will be discussed in detail in the following sections.

Observation model design
SR is a reverse engineering process, where we attempt to reconstruct the unknown HR images from

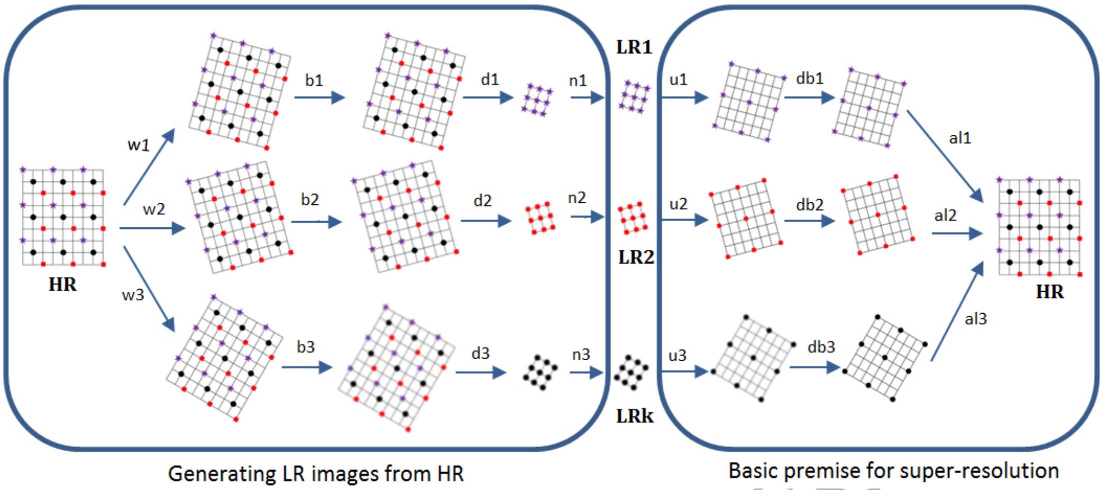


Figure 2: Generating LR images from a HR image and a basic approach of reverse engineering super-resolution to reconstruct the original HR image from multiple LR images. The original HR image is warped (w_k), blurred (b_k), down-sampled (d_k) and noise added (n_k) to generate the LR images. A basic super-resolution approach will upsample, de-blur, align, then combine the LR images to reconstruct the HR image.

one or more LR images of the same scene. To understand the intrinsic process of generating LR images from a HR image, the relationship between HR images and LR images has to be formulated in the form of an observation model. In the literature, several models have been proposed, but the most generally accepted observation model is illustrated in Figure 2.

The scene, represented in terms of a HR image, is captured multiple times by a camera or by multiple cameras to generate the low resolution we observe. The original high-resolution image is warped, blurred, down-sampled and noise-added to generate the corresponding LR images,

$$Y_k = D_k(B_k(W_k(X))) + N_k, \tag{1}$$

or in the matrix form,

$$Y_k = H_k X + N_k = D_k B_k W_k X + N_k, \tag{2}$$

where Y_k denotes LR images, D_k is a sub-sampling matrix, B_k is the blur matrix, W_k is the warp matrix, X is the original HR image, and N_k is the additive noise that corrupts the image. Most techniques proposed for the reconstruction of a HR image from LR images are based on the above observation model.

Estimating X from Y_k is an ill-posed problem and the results can differ vastly depending on the methods adopted to solve this problem. Considering the aforementioned observation model, a basic solution for the super-resolution problem is illustrated in Figure 2. The LR images are up-sampled, de-blurred, aligned and fused to generate the estimated HR image. One key to the success of this reverse-engineering process is the accuracy of the alignment/registration as discussed in the following section. Notice that this observation model is only applicable to the multiple input images case.

Image registration

Image registration is the process of aligning two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors by geometrically aligning the images onto a common reference grid. Registering images involves defining a mapping or a transformation for pixels from the sensed to the reference image. The transformation can be modeled by the following formula,

$$I_2(x, y) = g[I_1(f(x, y))], \tag{3}$$

where $I_1(x, y), I_2(x, y)$ are the pixel values at coordinates (x, y) in images I_1 and I_2, f is a transformation that maps the spatial coordinates and g transforms the intensity. The registration process usually consists of the three steps: feature detection and matching, transformation parameters estimation, and warping. There are two classes of transformation - global and local [26]. Global methods use the same transform parameters for the whole image while local methods treat individual regions differently. Global transformations are useful when the scene is relatively static, while local transformations are suitable when objects in the scene move and change independently.

Reconstruction
The final step of a super-resolution approach following registration is reconstructing the HR from the aligned sequence generated from the previous step. There are two categories of the final step: reconstruction-based and learning-based. While the reconstruction-based approaches try to *recover* the lost high-frequency components, the learning-based approaches try to *learn* them instead. The reconstruction-based approaches seek the lost high-frequency components in multiple complementary LR images. On the other hand, the learning-based approaches seek them from the training datasets. Reconstruction-based methods operate directly on the pixel values of the low-resolution images without prior knowledge, so these methods can be applied to images of various subjects. These algorithms can be divided into two classes: frequency domain and spatial domain [26]. While reconstruction-based super-resolution methods try to recover lost high-frequency components caused by aliasing, learning-based methods synthesize them instead [26]. A set of high-resolution images and corresponding low-resolution image patches are used to train the system by providing prior knowledge to the reconstruction process. These methods almost always produce visually pleasing images due to the high-frequency components created by the process. The problem is that although the reconstruction error is high, the resulting super-resolution image is often still a clear image, but it is in reality different to the original HR image as the high-frequency components that are inserted may not be appropriate for the image being resolved. Since learning-based super-resolution techniques may introduce spurious high frequencies, these approaches find it challenging to improve the performance of recognition based on super-resolved images.

3. Super-resolution for biometrics

While the first SR paper dated back to 1974 [29], the first SR work proposed for a biometric modality was presented 11 years later by Mjolsness in 1985 in [59]. In his Ph.D. thesis, Mjolsness proposed a SR approach for fingerprints. Subsequently, the first SR work proposed for the face was presented in 2000 by Baker and Kanade in [60]. Three years later, SR was introduced to the iris in 2003 by Huang *et al.* [44]. The first SR approach for gait was proposed in 2005 in [61] by Shechtman *et al.* when they first discussed space-time super-resolution. The timeline of landmark papers on super-resolution for biometrics is illustrated in Figure 3.

It has to be stated clearly here that the differences between applying SR in biometric applications and in general applications are not only the statistics of the input images, the object structures

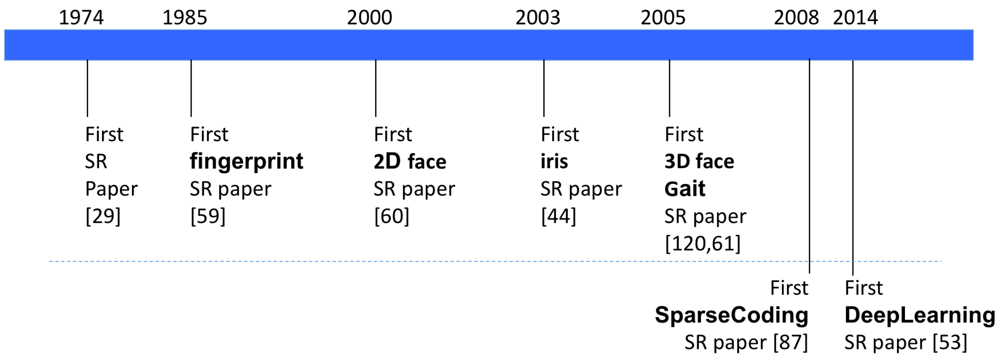


Figure 3: Timeline of the landmark papers on super-resolution for biometrics.

on the images, but mainly the expected output. *While general SR approaches work to improve the visual effect and clarity for human visual perception of the output HR images, biometrics SR approaches seek to improve the recognition performance, not the visual appearance.* While aiming to improve the recognition performance, SR approaches applied for biometrics not only impose specific prior knowledge and constraints on the objective functions and models, but also take advantages of special design of the biometric recognition and classification process to improve their performance.

In the following subsections, super-resolution techniques applied for multiple biometric modalities (face (2D + 3D), iris, gait and others) are reviewed.

3.1. Super-resolution for 2D face

A plethora of work has reported the adverse impact of low image resolution on face recognition systems [62, 63, 64, 65, 66]. Lemieux and Parizeau [62] reported significant decline in face recognition performance based on an eigenface (Principal Component Analysis - PCA) approach when the distance between two eyes was reduced to around 7 pixels (threshold). Similarly, Wang *et al.* [63] found a threshold of 32 pixels and Boom *et al.* [64] found a threshold of 16 pixels for a dramatic recognition performance degradation for both PCA and Linear Discriminant Analysis (LDA) recognition approaches. Fookes *et al.* [66] and Lin [65] tested the face recognition performance based on PCA and Elastic Bunch Graph Matching approaches with different eye distances of 111, 56, 28, 14 and 7 pixels. They observed a severe drop of recognition performance on both approaches when the resolution (eye distance) decreases. Empirical studies in [67] showed that minimum face resolution between 32×32 and 64×64 is required for existing algorithms. These papers have proved that the minimal resolution of recognition degradation varies differently for different recognition approaches.

Despite the adverse impact on the recognition performance of face systems, the low resolution is, in many cases, inevitable due to the limited acquisition conditions. To deal with the resolution reduction, Hennings-Yeomans *et al.* [68] presented three approaches for matching a low-resolution probe face image to high-resolution gallery images as shown in Figure 4.

- Down-sampling: The gallery face images are down-sampled to a lower resolution, which is equivalent to the probe resolution, then the typical face recognition approaches can be employed.

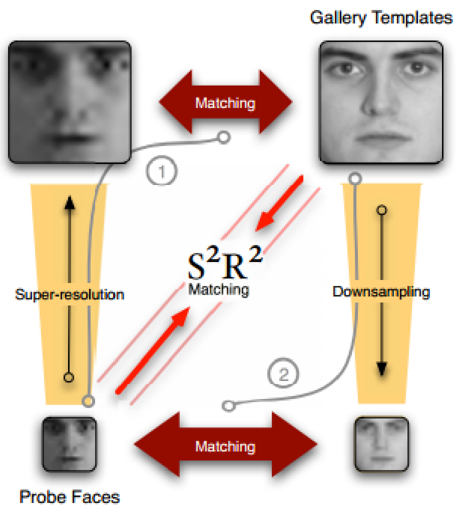


Figure 4: Three approaches for matching a low resolution probe face image to high-resolution gallery images: (1) Super-resolution ; (2) Down-sampling; (3) Simultaneous Super-Resolution and Feature Extraction [68]. Reprinted with permission from the authors.

- Super-resolution: The probe face image is super-resolved to a higher resolution, which is equivalent to the gallery resolution, then the typical face recognition approaches can be employed.
- Simultaneous Super-Resolution and Feature Extraction: These approaches integrate the tasks of super-resolution and recognition in one process either by directly computing a maximum likelihood identity parameter vector in high-resolution tensor space for recognition [69] or treating face features (e.g. Eigenfaces, Fisherfaces) as prior information in a super-resolution method [68].

While down-sampling is not in favor for face recognition since it decreases the amount of available information, super-resolution has been long used for face recognition and is gaining popularity due to the recognition improvement of the super-resolved images without the need to change the existing face recognition infrastructure. Simultaneously performing super-resolution and feature-extraction is considered as an extension of super-resolution, especially for the super-resolution in biometrics since jointly learning super-resolution and feature-extraction can lead to better recognition performance but this approach is still very daunting and not-universal when the approach needs to be specially designed for each chosen feature type. In this paper, we will discuss both super-resolution techniques and simultaneous super-resolution and feature extraction for the biometric applications.

One of the pioneering papers on super-resolution for the face is presented by Baker and Kanade [21]. High-frequency details of the images are inferred with probabilistic models from training samples. Followed by Baker and Kanade's work, Liu *et al.* proposed a two-step face hallucination approach by integrating a global parametric model with Gaussian assumption and a local non-parametric model based on Markov random field (MRF) [70]. From these two pioneering works, many super-resolution approaches have been proposed for facial images and it is still an active research topic now. In this section, we will approach the literature from three main perspectives: pixel-domain, feature-domain and deep learning-based.

3.1.1. Pixel-domain super-resolution for face

Given a set of training LR-HR image pairs, $S = (I_H^{(i)}, I_L^{(i)}), i = 1, 2, ..., N$ and an input query LR image, I_L, the problem is to estimate a corresponding HR image, $\widetilde{I}_H$, which is as close as possible to the original HR image.

MAP-based approaches

These approaches reconstruct a HR image from the given LR image by maximizing the conditional probability $P(I_H|I_L)$ as follows,

$$\begin{aligned} \widetilde{I}_H &= argmax_{I_H} P(I_H|I_L) \\ &= argmax_{I_H} \frac{P(I_L|I_H)P(I_H)}{P(I_L)}. \end{aligned} \tag{4}$$

Since the denominator, $P(I_L)$, is independent of the estimated output, I_H, the optimization can be reduced to maximizing the product of the likelihood, $P(I_L|I_H)$, and the prior, $P(I_H)$,

$$\widetilde{I}_H = argmax_{I_H} P(I_L|I_H)P(I_H). \tag{5}$$

The likelihood, $P(I_L|I_H)$, is generally modeled as a multivariate Gaussian distribution [21, 71, 72]. Since the relationship between LR and HR image pairs satisfies $I_L = HI_H$ where $H = DB_kW_k$ is the degradation matrix as shown in the previous section, the likelihood, $P(I_L|I_H)$, is modeled to minimize the error, $\|HI_H - I_L\|$.

Zhang and Cham [73, 74] chose to perform SR in the frequency domain of the Discrete Cosine Transform (DCT). Observing that the traditional interpolation approaches often lose the high-frequency components, the authors transform the LR input image into the DCT domain, estimating the Direct Current (DC) and Alternating Current(AC) coefficients, transforming the inferred DCT coefficients to generate a pre-filtered HR image. Some post-filtering steps can be applied to enhance the output. In the DCT domain, DC coefficients, which represent the average energy of a target block, can be estimated fairly accurately by simple interpolation-based methods (e.g. Bilinear, Cubic B-Spline), while AC coefficients, which contain the information of local features such as edges and corners around eyes and mouths of face image, cannot be estimated well by interpolation. They proposed an inference model to estimate AC coefficients by maximizing the posterior probability, $p(I_H^{AC}|I_L^{AC})$, where I_H^{AC} and I_L^{AC} are the AC coefficients of the HR and LR images respectively. The optimization problem can be formulated by a Markov network of a low-level vision field.

In case there are multiple LR images available, the optimization is represented as,

$$\widetilde{I}_H = argmax_{I_H} P(I_L^{(1)}, I_L^{(2)}, ..., I_L^{(M)}|I_H)P(I_H), \tag{6}$$

where $I_L^{(i)}$ is the i^{th} LR observation of the same HR image I_H. Gunturk [49] and Nguyen *et al.* [23, 52] showed that when the assumption of Independent Identical Distribution (IID) is valid, the joint conditional probability can be equal to the product of each individual conditional probability as follows,

$$P(I_L^{(1)}, I_L^{(2)}, ..., I_L^{(M)}|I_H) = \prod^{i} P(I_L^{(i)}|I_H). \tag{7}$$

Example-based approaches

Example-based super-resolution uses the known characteristics of this domain (i.e. the prior distribution) to perform specialized enhancement. They learn the priors from a database of HR images from the same domain (this is in contrast to priors defined by hand). Statistical pattern recognition methods are then used for example-based super-resolution. With the strong priors learned from the HR training dataset, example-based approaches are perfectly suited to face super-resolution. The example-based approaches construct the HR image as a linear combination of the HR training images as follows,

$$\widetilde{I}_H = \sum_i \widetilde{w}_i I_H^i, \tag{8}$$

where I_H^i are the training HR images and $\widetilde{w}_i$ are the corresponding weights. The weights are unknown and can be determined by minimizing the low resolution representation error as follows,

$$\widetilde{w}_i = argmin_{\{w_i\}} \left\| I_L - \sum_i I_L^i w_i \right\| = argmin_{\{w_i\}} E_L(w). \tag{9}$$

Ideally the low resolution representation error reaches a value of 0. In reality, there is a small representation error, $E_L(w) \leq \epsilon$. However, when this error is negligibly small, the solution for Equation 9 can be found by the following matrix form,

$$I_L = \sum_i I_L^i w_i = Iw. \tag{10}$$

According to the characteristics of a constrained least squares problem, Park *et al.* [75] proposed the optimal solution is obtained by the following equation,

$$\widetilde{w} = (I^T \cdot I)^{-1} \cdot I^T \cdot I_L. \tag{11}$$

Once the weights have been estimated, the reconstructed HR image is a simple linear combination of the HR training images with the estimated weights. To refine the results, the authors proposed a second step to employ recursive error back-projection to compensate for residual errors [75, 76].

Later on, Freeman *et al.* [77, 78] showed that local image information alone was normally insufficient to predict the corresponding HR patch and the spatial neighborhood effects must be taken into account. They used a MRF network to probabilistically model the relationship between HR and LR patches and between neighboring HR patches. The Markov network probability of any given HR patch choice for each node is given by,

$$P(x|y) = \frac{1}{Z} \Pi_{(ij)} \psi_{ij}(x_i, x_j) \Pi_i \phi_i(x_i, y_i), \tag{12}$$

where the transition function, $\psi_{ij}(x_i, x_j)$, represents the geometry compatibility of two adjacent HR patches (x_i, x_j) and the observation function, $\phi_i(x_i, y_i)$, represents the likelihood that the HR patch, x_i, matches the observed LR patch, y_i. Ignoring $\psi_{ij}(x_i, x_j)$ reduces the problem into the basic example-based SR presented above. The authors proposed to sample the input images' patches overlappingly. In the overlapped region, the pixel values of compatible patches should agree. Hence the likelihood of two adjacent HR patches $\psi_{ij}(x_i, x_j)$ can be measured inversely proportional to the difference between HR patches in the overlapping regions,

$$\psi_{ij}(x_i, x_j) = exp(-\frac{d_{ij}(x_i, x_j)^2}{2\sigma^2}), \tag{13}$$

where σ is a normalization parameter. The optimal HR patches are chosen to maximize the Markov network probability. However, finding the solution to this optimization may be intractable, the authors proposed an approximate solution and achieved promising results. This approach works with both general scenes and faces. The output super-resolved faces are significantly improved in sharpness in comparison with traditional bicubic interpolation.

Stephenson *et al.* proposed to use even stronger prior information by extending MRF-based super-resolution to use adaptive observation and transition functions, that is, to make these functions region-dependent [79]. Observing that the MRF model presented above is a type of Graphical Model (GM), the authors brought in some ideas of GM to modify the transition function, $\psi_{ij}(x_i, x_j)$ and the observation function, $\phi_i(x_i, y_i)$. Firstly, the search for a matching HR patch is restricted to a local region rather than the whole face image. Adapting the observation function, $\phi_i(x_i, y_i)$, to include region prior constraints results in better modeling the relationship between HR and LR patches. Secondly, observing that many of the HR patches in a face image may be strongly correlated with patches a long distance away since human faces are highly symmetrical, features found on one side of the face will typically be found on the other side, the normalization parameters in Equation 13 can be modified to reflect the long-distance neighborhood. Adapting two functions to restrict the likely high-frequency patches available for the super-resolution not only reduces the Mean Square Error (MSE) associated with a standard MRF but how using such adaptation can produce sharper images. While the local priors can be used either identically in all locations as in the spatially invariant local models or differently in different locations as in the spatially variant local models [80], global priors provide a balance between computation complexity and representation power [49, 81, 80]. Akyol *et al.* proposed to consider shape reconstruction as a separate problem and solve it together with texture reconstruction in a coordinated manner to both reduce the computation and improve the reconstruction accuracy [80].

Interestingly, a number of works have been suggested to estimate the weights in the example-based SR approach in other domains. Wang and Tang [82] applied PCA to fit the input LR face image as a linear combination of the low-resolution face images in the training set. The HR image is rendered by replacing the LR training images with high-resolution counterparts, while retaining the same combination coefficients. Transforming the representation into the PCA domain takes advantages of the structural similarity of face images. The super-resolved faces not only achieve enhanced visualization for human perception, but also make the automatic recognition procedure easier since they emphasize the face difference by adding more high-frequency details. The major disadvantage of this approach is that it only operates on global estimation without dealing with the local details, which may lead to a lack of detailed features and distortion of the output. However, Huang *et al.* argued that there are substantial differences between the manifolds of PCA coefficients for LR and HR face images [83]. They proposed using canonical correlation analysis (CCA) to determine a coherent subspace in which the statistical correlation between intrinsic structures of LR and HR images is maximized. Using CCA separately for reconstruction of global face appearance and facial details, they achieved improved reconstruction quality over existing state-of-the-art super-resolution algorithms, both visually, and using a quantitative peak signal-to-noise ratio assessment. Differently, Liu *et al.* [84] choose to perform SR in the Tensor domain. Each patch in the LR input image is represented as a multilinear combination of training LR patches in the TensorPatch domain. The HR patch can be synthesized by applying the weights inferred from input LR patches. To further enhance the quality of the hallucinated image, a coupled residue compensation algorithm is applied based on a new statistical index called Coupled PCA, which infers the residual from LR residue to HR residue.

Sparse Representation-based approaches
Most recently, inspired by the success of sparse representation in face recognition [85] and super-resolution [86, 87], researchers have proposed to perform super-resolution for face images based on sparse coding theory [88, 89, 90, 91, 92, 93, 94, 95]. Denote two matrices, X_H and X_L, as the HR and LR training matrices where each column is the HR or LR training vector respectively. Sparse coding theory states that a vector can be represented as a sparse linear combination of base vectors in the over-complete dictionary [87]. The problem of finding a sparse solution c of a LR input image, I_L (not included in the training set), over the corresponding given over-complete dictionary can be defined as follows,

$$\tilde{c} = argmin_c \left\| c \right\|_1, s.t. \left\| I_L - X_L c \right\|_2^2 \leq \epsilon. \tag{14}$$

Lagrange multipliers offer an equivalent form as,

$$\tilde{c} = argmin_c \left\| I_L - X_L c \right\|_2^2 + \lambda \left\| c \right\|_1. \tag{15}$$

The corresponding HR image, I_H, can be simply reconstructed as follows,

$$\tilde{I}_H = X_H \tilde{c}. \tag{16}$$

This fundamental idea for general super-resolution [87] has been applied for reconstructing separate patches of faces [88, 96].

Observing that sparse coding does not ensure a good locality and may fail to facilitate the nonlinear function, Jiang *et al.* [97, 98] proposed to incorporate prior support information induced by locality to the minimization in Equation 14 with weighted minimization,

$$\tilde{c} = argmin_c \left\| c \right\|_{1,a}, s.t. \left\| I_L - X_L c \right\|_2^2 \leq \epsilon, \tag{17}$$

where $\left\| c \right\|_{1,a}$ is the weighted L_1 norm, which is defined as,

$$\left\| c \right\|_{1,a} = \sum_{i=1}^{N} \alpha_i \left| c_i \right|, with \alpha_i = \begin{cases} 1 \; i \in T \\ \infty \; otherwise. \end{cases} \tag{18}$$

T is the support information defined as: $T = supp(dist|_k)$, where $dist|k$ is k-nearest bases of the measurements of distances between I_L and the bases in the LR dictionary. What the support information does is to limit the varying region of the weight, forcing the weight to be estimated locally. The sparse weight estimation with the locality constraint enhances accuracy and stability.

The sparse solution c intuitively represents an embedding geometry of the image space. There is one assumption here is that the embedding geometry of the HR image space, c_H, and its counterpart of the LR image space, c_L, are identical and represented as c. In [89, 90], Gao *et al.* experimented and proved that this assumption is sometimes incorrect, and may contribute to the unsatisfactory reconstruction results of a SR approach. Some authors train the HR and LR dictionaries jointly to make two sparse solutions consistent with each other [99, 100]. Observing that the relationship between two sparse solutions can be learned, Gao *et al.* introduced a mapping function, P, between c_H and c_L as,

$$\Delta_H = P\Delta_L, \tag{19}$$

where $\Delta_H = [c_H^{(1)}, c_H^{(2)}, ...]$ and $\Delta_L = [c_L^{(1)}, c_L^{(2)}, ...]$. The mapping function, P, collapses to an identity matrix if the assumption of identical HR sparse solution, c_H, and LR sparse solution, c_L, holds [91, 82, 87, 101]. The mapping function is learned by solving the following regression problem,

$$\hat{P} = argmin_P\{\|\Delta_H - P\Delta_L\|_F^2 + \gamma\,\|P\|_F^2\}. \tag{20}$$

Once the mapping function is learned, the estimated HR image is reconstructed as $\tilde{I}_H = X_H P\tilde{c}$ rather than Equation 16. To obtain the HR images with higher fidelity, the reconstructed results are further treated as a preliminary estimation for the target HR image in the MAP estimation,

$$\hat{I}_H = argmin_{I_H}\{\|DBI_H - I_L\|_2^2 + c\left\|I_H - \tilde{I}_H\right\|_2^2\}, \tag{21}$$

where the image degradation model is incorporated in the form of the down-sampling matrix, D, and the blurring matrix, B. The estimation can be achieved by iterative gradient descent. The advantages of this approach are the separation of the embedding geometries in the HR and LR image spaces and the use of sparse representation as a preliminary estimation for the super-resolution MAP problem.

While other approaches represent image patches as a sparse linear combination of patches from an appropriately chosen over-complete dictionary, Li *et al.* proposed to represent the central pixel as a sparse linear combination of neighboring pixels. Observing that the HR face and its corresponding LR one have both similar global face structure and local-pixel structure, the authors introduced a three-step framework for the face hallucination problem. In the first step, the input LR face is used to search the face dataset and identify the k pairs of LR-HR example faces having the most similar local-pixel structures to the input LR face, using PCA and k Nearest Neighbor (kNN). In the second step, the local-pixel structures, which are represented by the weights of the neighboring pixels, are learned from the k HR example faces using sparse representation. The final step is to estimate the target HR face iteratively using the weights of the neighbors.

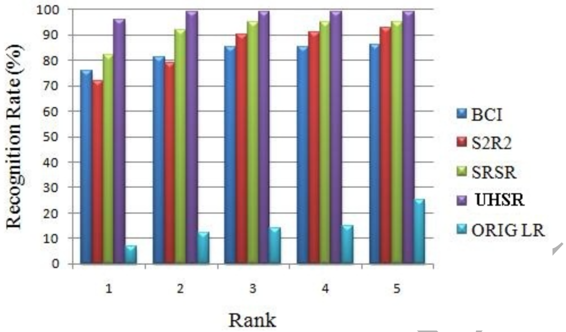
Bilgazyev *et al.* tackled the sparse representation-based super-resolution for LR face recognition at a distance in the frequency domain [92]. The HR and LR face images are transformed into frequency domain by Dual Tree Complex Wavelet Transform (DT-CWT). DT-CWT extracts high frequency components from a database of HR facial images and synthetically generated LR images to generate the high frequency HR and LR dictionaries, D_x^Ψ and D_y^Ψ, respectively. Then the super-resolution operation is performed in the generated high frequency domain. Given a LR input image, the sparse representation of the high frequency components of this LR image in the high frequency LR dictionary is computed as follows,

$$\alpha_i^* = argmin_i \left\|\vec{\theta}^{\Psi}{}_{y,i} - D_y^\Psi \alpha_i\right\|_2 + \lambda\,\|\alpha_i\|_1\,, \tag{22}$$

where $\vec{\theta}^{\Psi}{}_{y,i}$ is the high frequency component of the given LR images, and λ is the weight on sparsity. Once the sparse representation of the given LE image, α_i^*, has been computed, the high frequency components of it can be estimated as follows,

$$\hat{\theta}_{x,i}^\Psi = D_x^\Psi \alpha_i^*. \tag{23}$$

(a) LR input image from a surveillance camera and SR output image



(b) The rank-one recognition rates of face recognition on a surveillance dataset [92]

Figure 5: Face recognition performance of some sparse-representation super-resolution approaches: BCI (Bicubic), S2R2 (Simultaneous SR and Recognition [68]), SRSR (SR via Sparse Representation [87]), UHSR (University of Houston SR [92])). Reprinted with permission from the authors.

The corresponding super-resolved HR image can be estimated as follows,

$$\hat{X} = \Psi^{-1}(\Psi L^{-1} Y + \hat{\Theta}^{\Psi}), \hat{\Theta}^{\Psi} = \{\hat{\theta}^{\Psi}_{x,1} ... \hat{\theta}^{\Psi}_{x,2}\}, \tag{24}$$

where L^{-1} is the up-sampling operator. Hence by estimating the high frequency components, rather than studying the direct relationship between the HR and LR images, this approach achieved impressive reported recognition performance for both synthetic and surveillance facial databases.

Some authors proposed to work on constructing dictionaries expressively [99, 100, 102]. They first obtained a couple of HR and LR over-complete dictionaries using patches in the same position as the input LR face image from all training samples. Subsequently, they refined the dictionaries with Locality Preserving Projections (LPP) based K nearest neighboring position-patch selection. The patches in LPP transformed space have better distinguish-ability, which leads to more expressive dictionaries.

Face recognition performance of some sparse-representation super-resolution approaches is presented in Figure 5.

3.1.2. Feature-domain super-resolution for face (Recognition-oriented SR for face biometrics)

Observing that most face recognition systems employ a pre-processing step to reduce the dimension of the image prior to further processing, Gunturk *et al.* proposed to transfer the SR reconstruction from the pixel domain to the lower dimensional face space. The original image, x, is transformed to low resolution image by Down-sampling, $D^{(i)}$; Blurring, $B^{(i)}$; and Warping, $W^{(i)}$, as follows,

$$y^{(i)} = D^{(i)} B^{(i)} W^{(i)} x + n^{(i)}, \tag{25}$$

where $y^{(i)}$ is the low resolution image and $n^{(i)}$ is the observation noise. The original and low resolution images are represented as a combination of coefficients in the eigenface domain,

$$x = \Phi a + e_x, \tag{26}$$

$$y^{(i)} = \Psi \hat{a}^{(i)} + e_y^{(i)}, \tag{27}$$

where e_x is the error vector and a is the feature vector of the original HR image; $e_y^{(i)}$ is the error vector and $\hat{a}^{(i)}$ is the feature vector of the ith LR image. Φ and Ψ are the eigenvector base matrices of the HR and LR images. Substituting Equation 26 and 27 into Equation 25, we have,

$$\Psi \hat{a}^{(i)} + e_y^{(i)} = H^{(i)} \Phi a + H^{(i)} e_x + n^{(i)}, \tag{28}$$

where $H^{(i)}$ incorporates the motion, blurring, and downsampling processes as $H^{(i)} = D^{(i)} B^{(i)} W^{(i)}$. Withdrawing $a^{(i)}$, we have,

$$\hat{a}^{(i)} = \Psi^T H^{(i)} \Phi a + \Psi^T H^{(i)} e_x + \Psi^T n^{(i)}. \tag{29}$$

Equation 29 presents the relationship between the high-resolution features and low resolution features. The relationship between high-resolution and low resolution image of Equation 25 has been transformed into the feature domain. From this feature-domain relationship, the high-resolution features can be estimated using a maximum a posterior (MAP) approach to maximize the product of the conditional probability, $p(\hat{a}^{(1)}, ..., \hat{a}^{(M)}|a)$ and the prior probability, $p(a)$,

$$\tilde{a} = argmax_a p(\hat{a}^{(1)}, ..., \hat{a}^{(M)}|a) p(a). \tag{30}$$

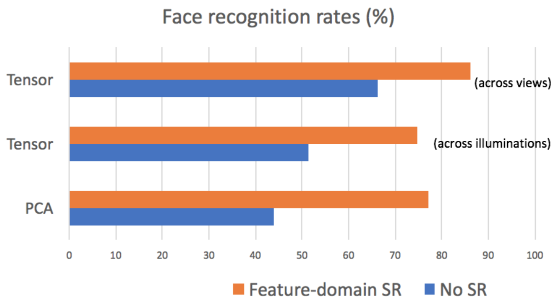
These approaches no longer super-resolve images in the pixel-domain, but super-resolve the extracted features that are used for classification in the feature-domain, and the SR output (a super-resolved feature vector) is directly employed for recognition. Different linear features including PCA [49], Tensor Face [69, 50] and Gabor-based [23, 52] have been investigated to improve biometric performance. These features are super-resolved using a MAP estimation approach. Specific knowledge of face models is incorporated in the form of prior probabilities to constrain the SR process, improving robustness to noise and segmentation errors. These approaches have been shown to outperform the equivalent pixel-domain SR approaches for face recognition.

Face recognition performance of some feature-domain super-resolution approaches is presented in Figure 6.

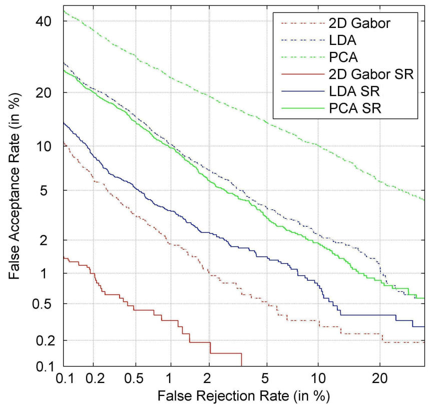
3.1.3. Deep learning-based approaches

Deep learning has recently revolutionized many computer vision tasks, including object detection, object segmentation, object recognition, object tracking, action recognition and scene understanding [103, 104, 105]. Deep learning techniques are designed to simulate functions of the human brain in a hierarchical multiple-layer architecture. The major advance of these techniques is the capability to learn to represent features by themselves based on the nature of the data, rather than the subjective nature of human perception [104] in an end-to-end learning approach. Recently, deep learning has been introduced to super-resolution and has shown promising reconstruction accuracy.

The very first work of deep learning in super-resolution was presented by Dong *et al.* in 2014. In [53], the authors presented a neural network named Super-Resolution Convolutional Neural Network (SRCNN) to directly learn an end-to-end mapping between LR-HR images. The deep network has three layers, with each layer simulating various steps in the sparse-coding super-resolution approaches. The authors argued that while sparse-coding super-resolution approaches pay particular attention to learning and optimizing the dictionaries [87], the rest of the pipeline is rarely optimized or considered in a unified optimization framework. They showed that the



(a) PCA [49] and Tensor [69] features



(b) PCA, LDA, Gabor-based features [23]

Figure 6: Face recognition performance of some feature-domain super-resolution approaches. Reprinted with permission from the authors.

sparse-coding super-resolution pipeline could be equivalent to a deep convolutional neural network. Simulating the pipeline by a deep Convolutional Neural Network (CNN) allows the network to learn LR-HR mapping in an end-to-end manner, with hidden layers trained to learn the dictionaries implicitly [53]. This network achieved state-of-the-art reconstruction accuracy in natural image datasets [53, 106].

With the same idea of simulating the sparse-coding super-resolution approaches with deep neural networks, Wang *et al.* simulated the sparse coding algorithm by a Learned Iterative Shrinkage and Thresholding Algorithm (LISTA) [107, 108]. A multi-layer neural network mimicking the sparse-coding super-resolution is then proposed and trained end-to-end to learn the mapping between LR and HR images. With the domain expertise of the conventional sparse coding approach, the proposed sparse-coding super-resolution (SCN) outperforms SRCNN.

Interestingly, some authors discuss the involvement of context in the super-resolution process. Kim *et al.* argued that if the network can look at a larger context to understand the surrounding neighbors, it can reconstruct the HR counterpart more accurately [56]. With that observation, they proposed a very deep convolutional network (20 layers) inspired by VGG-net to cover larger image context [56]. To speed up the convergence, they proposed to learn residuals only and used extremely high learning rates. This very deep super-resolution (VDSR) approach achieved superior results in comparison with SRCNN. To reduce the number of parameters and the likelihood of over-fitting, some authors proposed to share the parameters between middle layers, which means recursively applying the same convolutional layers multiple times [109, 110]. Also to incorporate stronger context, either a multiple scale network (Laplacian Pyramid Networks [111]) or an ensemble of multiple networks [58] can be used to progressively reconstruct the high-resolution images.

There are two inherent limitations restricting the running speed of the existing works: the up-sampling pre-processing of the original LR to the desired size and the expensive non-linear mapping. To solve these limitations, some authors proposed to shift the up-sampling to the end of the network with a deconvolution layer [112] or with a sub-pixel convolutional layer [55] to reduce the computation complexity. Both methods achieved real-time performance either with Graphical Processing Unit (GPU) [55] or with a generic Central Processing Unit (CPU) [112] still achieving superior reconstruction accuracy.

It is also noteworthy that recent Generative Adversarial Networks (GAN) with its super performance in generating images has also been incorporated into the super-resolution pipeline to learn to generate a HR image from the LR counterpart. This SRGAN approach achieves photo-realistic performance and is the latest state-of-the-art in super-resolution from a single image now [57].

With their superior performance, the aforementioned state-of-the-art super-resolution approaches based on deep learning have been applied to the face recognition task to improve recognition performance. These generic CNN-based approaches can be applied directly to the facial images to generate a HR image as a pre-processing step before feeding the HR output image to the recognition phase such as in [113] to achieve 6 to 10% improvement in the recognition rate. However, applying generic SR approaches directly to facial images is a naive process since the prior on face structure or face spatial configuration has been proved to be pivotal for the face super-resolution task [72]. Berger *et al.* further refined the HR output with a localized SR step [114]. The localized SR step focused on locally reconstructing image patches at crucial face landmark points (e.g. eyes, nose, mouth) via dictionary learning. However, having two networks in cascade is computational expensive. Ko *et al.* also employed multi-task neural networks for various landmark points, but shared lower layers among these networks to reduce the computation [115]. The bi-channel CNN applied by Zhou *et al.* [116] shared the same idea of multi-task (in this case is two-task) neural networks, where the lower layers are shared, the upper fully connected layers are divided into two groups. One group predicts a reconstructed face image, I_{rec} and the other group estimates a fusion coefficient. The final HR output image is a linear combination of two channels: the reconstructed face image, I_{rec} and the LR input image. Differently, Zhu *et al.* extended the notion of prior to pixel-wise dense face correspondence field [117].

The super-resolution and dense correspondence field estimation tasks are jointly optimized in a deep cascaded bi-network. The network consists of a common branch, a high-frequency branch and a gated network which is trained and learned jointly. The major advantage of this deep cascaded bi-network is the capability to work with very LR input images (several pixels inter-ocular distance measured pxIOD). Recently, Deep Reinforcement Learning (DRL) has also been used to super-resolve face images [118]. Different from traditional deep learning approaches, DRL provides a policy learning to navigate the attention to different regions to perform super-resolution, which is capable of adaptively personalizing an optimal searching path for each face according to its own characteristic. Samples of the super-resolved faces of some deep-learning based approaches are illustrated in Figure 7.

To summarize, each category discussed tackles the problem from a different point of view. While MAP-based approaches provide a nice intuition from a probabilistic point of view, the example-based and sparse-representation-based approaches provide intuitive insights from a reconstruction point of view. The major drawback of the MAP-based approaches is the assumption about IID, which may not be valid, especially in the video cases. In contrast, the example-based and sparse-representation-based approaches have a challenging limitation in learning and designing the optimal dictionary. The deep-learning-based approaches have shown great potential in learning and modeling the complicated relationship between LR images and their HR counterparts, even in an end-to-end manner. However, the major drawbacks of these deep approaches are the expensive computational complexity (usually requiring GPUs) and they are data-hungry since they require a huge amount of training data. Different from the above approaches, the feature-domain approaches shift

(a) Sample super-resolved output images

	LFW 4x	LFW 8x	BioID 4x	BioID 8x
Bicubic	26.79	21.92	25.18	20.68
SFH	26.59	22.12	25.41	20.31
BCCNN	26.60	22.62	24.77	21.40
MZQ	25.93	22.12	24.66	21.11
SRCNN	28.94	23.92	27.02	22.34
VDSR	29.25	24.12	28.52	24.31
GLN	30.34	24.51	29.13	24.76
ARL	32.93	26.17	31.56	26.56

(b) Peak Signal-to-Noise Ratio (PSNR)

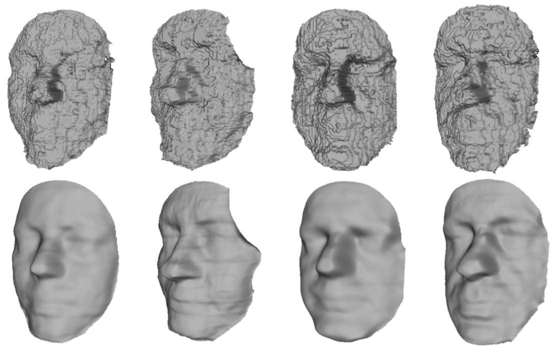
Figure 7: Samples of some deep-learning based super-resolution approaches on the LFW dataset: SRCNN [53], VDSR [56], GLN [119], ARL [118]. Image adapted from [118].

the super-resolution process to the feature domain, which is used directly by the recognition step. This leads to better constraints in the reconstruction, but also requires heavy expert knowledge on the feature domain of the researchers.

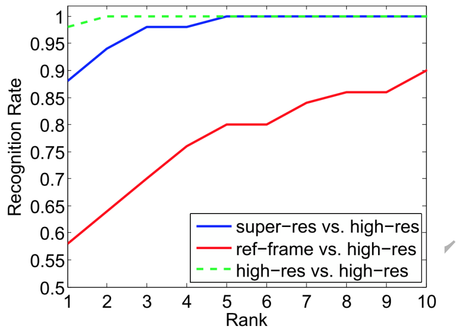
3.2. Super-resolution for 3D faces

Following the success in 2D face super-resolution, the research community has turned to super-resolve LR 3D face models into HR 3D models. Peng *et al.* employed a learning-based approach to learn the mapping between LR 3D models and HR 3D models [120]. The LR 3D model is regularized into a 2D regularized LR image by projecting it into a cylindrical coordinate surrounding the 3D face model. Laplacian pyramids are constructed for both 2D regularized LR images and 2D regularized HR images from the training data to construct the Parent Structure. The Parent Structure is a multiple level feature representation of the image. While the 2D regularized HR images are represented in the Parent Structure from level 0 to level N, the lower-resolution images are only represented as part of that, from level k to level N. The reconstruction task now becomes predicting the level 0 to $k-1$ of the regularized LR images based on the training set. An approach similar to [60] is then applied to predict the missing information from level 0 to level $k-1$. In a similar vein, Pan *et al.* replaced the Laplacian pyramid representation by the Progressive Resolution Chain (PRC) representation of 3D models [121]. The PRC represents a 3D model as a hierarchy of 3D models with different levels of resolution. The authors also replaced the cylindrical coordinate projection with a consistent planar representation. The HR 3D face model is restored by a MAP approach.

Berretti *et al.* employed a three-step approach to derive one HR 3D face model from several LR depth images acquired through a Microsoft® Kinect scanner [122, 123]. The first step is a face detector, which detects and crops the face region from each frame. These cropped faces are then fed into the registration module that aligns LR 3D models to the first one based on the scaled Iterative Closest Point (ICP) algorithm. Based on the statistics of the observation values, the third step is employed to re-sample the aligned LR 3D models at a higher resolution to reconstruct the final HR 3D model, which they called the SuperFace. Recently, Bondi *et al.* extended the ideas to deal with uncooperative contexts [124]. Rather than asking the subjects to sit in front of the Kinect camera to perform head movement, they allow the subject to pass in front of the scanner, following an uncooperative protocol. Non-rigid registration techniques called the Coherent Point Drift (CPD)



(a) LR 3D inputs and SR 3D outputs



(b) The recognition rates of 3D face recognition on a dataset [92]

Figure 8: One example of super-resolution for 3D face recognition performance [123].

is employed to register point sets to account for deformation of the face. Registered 3D data are filtered through a variant of the lowess method to remove outliers and build the final face model.

Ouji *et al.* proposed a 3D acquisition system using three calibrated cameras coupled with a non-calibrated projector device [125]. This combination is rapid, easily movable and robust to ambient lighting variation. Then the authors performed a 3D temporal super-resolution to correct the 3D information provided by the spatial super-resolution and to deal with 3D artifacts caused by either an expression variation, an occlusion or even a facial surface reflectance [125]. The non-rigid registration technique (CPD) is employed to register successive 3D frames and localize and clear the artifacts representing a high spatial deviation. Once registered, the 3D point sets and also their corresponding 2D texture images are used as LR data to create a HR 3D point set and its corresponding texture. One example of super-resolution for 3D face recognition performance is illustrated in Figure 8.

One body of work uses 3D models to assist the super-resolution process of 2D facial texture [126, 127, 128]. Yu *et al.* used a generic 3D face model in a reconstruction-based framework to estimate the HR texture image [126]. A LR video of a face with varying pose is the input for the algorithm. After estimating the initial pose and illumination from the first frame, the generic 3D model's texture is updated sequentially by error back-projection in each frame. Mortazavian *et al.* approached the estimation by using a 3D morphable model in an example-based framework for a single image input [127, 128]. The 3D morphable model is fitted into the input LR image. The fitted 3D morphable model and the estimated shape and rendering parameters are used to extract the facial texture from the LR input image. Once the facial texture is extracted, an example-based super-resolution approach is performed to generate a HR facial texture.

The accuracy of the super-resolved 3D models is evaluated based on two major methods:
(1) Visually - error map: for human perception, the accuracy of the super-resolution is presented in the form of an error map. The error map is constructed by first aligning the reconstructed 3D face model and the 3D HR face model, then for each point of the reconstructed model, its distance to the closest point in the 3D HR model is calculated. The error map comparably presents the level of error for each point and each section of the reconstructed 3D model.
(2) Quantitatively - Root Mean Square Error (RMSE) and Relative Mean Deviation (RMD). RMSE

quantitatively represents the difference between the reconstructed 3D model, S, and the 3D HR model, S', as the average error of each point pair,

$$RMSE(S,S') = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - p_i')^2}, \tag{31}$$

where p and p' are the closest points from model S and S' and N is the number of correspondence points between two models. Relative Mean Deviation (RMD) can be defined as the mean of the absolute deviations of reconstructed 3D model, S, and the 3D HR model, S',

$$RMD(S,S') = \frac{\sum_{i=1}^{N}|p_i - p_i'|}{N}. \tag{32}$$

Datasets: A number of authors down-sampled the high-resolution 3D face models to low resolution 3D face models. In the literature, the University of South Florida (USF) Human ID 3D face database [129] has been used by [120, 121]. This database consists of HR 3D models of 136 subjects. There are more than 90,000 vertices and 180,000 facets for each model. Another database which has been used in the literature is Superfaces Florence [122]. This database consists of one sequence of depth frames acquired through a Kinect scanner as well as one high-resolution face scan acquired through a 3dMD scanner per subject. The HR 3D mesh has 40,000 vertices, 80,000 facets and a texture stereo image with a resolution of 3341×2027 pixels. The Kinect scanner captured 10 to 15 seconds at 30 frames per second for each subject who was asked to move the face around the jaw axis up to an angle of 60 to 70 degrees for both left and right face side visibility. There are 50 subjects in this database.

3.3. Super-resolution for iris

The iris is one of the most robust and accurate biometric traits due to a high degree of randomness of its features, high permanence, and limited genetic penetrance [130, 131]. With its advantages, iris recognition has found popular deployment in real-life such as airports for access control or border security applications. Notwithstanding, most conventional iris recognition systems impose heavy constraints on the participants, including “stopping and staring” at a close distance (less than 60 cm) to the camera, which slows down the throughput and causes inconvenience to the participants. To deal with this issue, the recent trend is to extend the distance of image acquisition with less constraints imposed on the subjects, called “iris recognition *at a distance* and *on the move*” [3, 132, 133, 134, 135].

The major challenge for iris recognition *at a distance* and *on the move* is the small size of the eyes (the diameter of an eye is on the order of 1 cm). When the eyes are imaged at long distances, the resolution of the acquired images is normally low in comparison with the standard recommended resolution for a confidence matching of 120 pixels across the eye [136]. The lack of resolution, similar to the case of face recognition, has adverse impacts on the iris recognition performance [137, 135]. Super-resolution techniques have been proposed to increase the resolution of the iris images and improve the recognition performance of the iris biometrics systems. The first work recorded dated back to 2003 when Huang *et al.* proposed a learning-based super-resolution approach for the iris to improve the classification process [44]. Since then, a number of approaches have been introduced. This section reviews the state-of-the-art super-resolution techniques for iris images. We categorize these techniques based on the domain in which the super-resolution process is performed.

3.3.1. Pixel-domain super-resolution for iris

Learning-based SR

Huang *et al.* [44] proposed a learning-based method based on the Circular Symmetric Filter (CSF). Their algorithm predicted the prior relationship between iris feature information in different bands and incorporated it into the process of iris image enhancement. The test image is first scaled up by cubic interpolation, then filtered using the CSF to generate the initial features. The authors then broke the obtained medium frequency image in a raster-scan order, and normalized the low-resolution patches using a local contrast normalization method. The low-resolution patches are subsequently scanned over using the L1 distance to find 200 patch pairs from the training set whose feature vectors are closest to the input patch at each step. The best matching pair from this sub-set of patch pairs are selected based on spatial constraints at adjacent patch borders. The authors then reversed the normalization for the predicted high-resolution patch and add it to the corresponding output patch. Finally, the desired super-resolution image is achieved by adding the high-frequency image to the test image. Experiments shows improvement in correct classification rates for the proposed approach in comparison with the traditional interpolation super-resolution and the approach of [78].

Kwang *et al.* [45] proposed another learning-based super-resolution method based on multiple MLPs (multi-layer perceptrons). The middle- and high-frequency components of a low-resolution iris image are restored from the trained neural network architecture. The procedure is explained as follows: (1) Each LR image is divided into small blocks (size 4×4, or 5×5 depending on the original size of the LR image); (2) Each block is classified into one of 3 types: vertical, non-edge, and horizontal, based on the following criteria:

- Vertical if $(V - H) >= T$
- Non-edge if $-T <= (V - H) < T$
- Horizontal if $((V - H) < -T$

where $V = |(I_1 + I_3) - (I_2 + I_4)|$, $H = |(I_1 + I_2) - (I_3 + I_4)|$, and T is a threshold to determine edge direction; (3) These 3 types of blocks are fed into 3 corresponding multi-layer perceptron neural networks to estimate the selected pixel values ($O_1, .., O_4$ for the block size 4×4, and $O_1, ..., O_5$ for the block size 5×5); and (4) Blocks are reassembled into a HR iris image. Pixels without values are filled in using bi-linear interpolation. This proposed approach has been shown to outperform conventional bi-linear interpolation in terms of the Equal Error Rate (EER) of the iris recognition system.

Both Huang *et al.* [44] and Kwang *et al.* [45] reported good performance in visual and recognition improvement. However, the robustness of iris recognition is due to the high level of distinction among different irises; the learning process as used in [45, 44] can introduce spurious high frequencies, which may mislead a recognition procedure. In addition, both methods are conducted in artificially-created low-resolution images (low-resolution iris images are produced by degrading high-resolution-images with a Gaussian kernel and down-sampling), casting doubt as to whether they will work with real low-resolution images that suffer from additional challenges. This is also the question of how to effectively learn the dictionary for these techniques.

Interestingly, deep learning-based super-resolution has found its way into the iris recognition domain. Zhang *et al.* employed a deep-learning-based super-resolution approach named SRCNN [53, 106] in combination with Super-Resolution Forests (SRF) [138] to generate two HR output images of the LR normalized iris images [139]. The experiments on The Chinese Academy of Sciences Institute of Automation, CASIA-Iris-Mobile dataset show improvements on the EER with SRCNN leading to less EER in comparison with SRF.

Reconstruction-based SR

In [46], the authors proposed the first technique of implementing super-resolution for iris images. The proposed procedure is as follows: (1) Select the best image from a sequence of given iris images; (2) Interpolate the selected image with zeros in order to enlarge; (3) Break the interpolated image into smaller patches, and align each patch with the template; (4) Combine information about numerical value for each pixel from scene patches (using linear combination) and fill the holes on template patches with new value; and (5) Smooth the boundaries of patches. Even though, the proposed idea has not been experimented for validation, this pilot research motivated further exploration and investigation from the research community.

In 2006, Barnard *et al.* introduced a novel multi-lens imaging hardware system to capture multiple iris images. The authors then presented a technique to compute a high-resolution reconstructed image from an ensemble of low-resolution images containing sub-pixel level displacements. The reconstruction technique is performed as an inverse problem of the observation model in Equation 25 to minimize the reconstruction error. The point spread function, B, the warp function, W, and the down-sampling function, D, are estimated before the optimization is done by the well known method of conjugate gradients (CGLS). The reconstructed high-resolution iris images have been shown to consistently improve the recognition rate in terms of the Hamming distances. However, the precision of parameter estimation is the key to the success of this approach. To avoid these estimations, Deshpande *et al.* interpolated one LR image to a HR image, then iteratively back-projected the difference between other LR images with a down-sampled version of the HR image to minimize the error [140, 141]. The major question for this approach is the effective way to iteratively back-project the difference in the LR domain to update the HR image.

In 2007, Fahmy [40] proposed a reconstruction-based super-resolution technique to reconstruct higher resolution iris images from video sequences. From 3 seconds of video (90 to 100 frames), a set of 16 frames is selected and registered using a cross correlation model [142]. From these 16 aligned frames, a set of 9 frames is chosen for reconstructing one HR iris image (no selection criteria is provided in the paper). The auto-regressive signature value between the reference frame (the first frame in the chosen sequence) and each of the 8 remaining frames is calculated in 3 directions: vertical, horizontal, and diagonal. Then, one frame with the highest auto-regressive signature value in each direction is selected. These 3 selected frames are interleaved with the reference frame to generate one image at 4-times higher resolution. This process can be iterated to reconstruct a 16-times higher resolution frame. Experiments show improvement in terms of recognition performance for the 4 and 16 times resolution images in comparison with the original LR image. However, Fahmy's [40] approach uses the whole eye image for registration, which is potentially problematic due to iris dilation and contractibility properties. Furthermore, the situation will be worse in less constrained iris recognition applications. In addition, the frames employed for super-resolution have to be in focus, otherwise the performance of the reconstruction will significantly degrade. This constraint makes this approach impractical and thus unsuitable for real life applications.

Observing that different frames in an iris video sequence captured in changing conditions may exhibit different quality, Nguyen *et al.* proposed to incorporate the quality measures into the super-resolution process [41, 42]. They relied on the Dempster-Shafer theory to combine four quality factors (focus, off-angle, motion blur and illumination variation) into one unified quality score for each iris frame. The unified quality scores are then employed to weight the contribution of the frames to the super-resolution process. Othman *et al.* expanded this approach by proposing an alternative method to qualify the quality of an iris image [143, 144]. They calculated a local quality (LQ) measure for each local region of the iris image based on a Gaussian Mixture Model (GMM) estimation of the distribution of a clean iris texture. The local quality measures used for weighted fusion can deal with the non-uniform quality variation in the iris image. In the same vein, Hsieh *et al.* also exploited the quality measurements to weigh the contribution of LR observations [145]. However, different from previous approaches, they performed the super-resolution on local patches rather than the whole image. Performing super-resolution on local patches better deals with local deformation normally presented in the normalized iris images, but the drawback of the local approach is effective stitching of the reconstructed HR patches at boundaries.

Similar to face, sparse representation can also be used to reconstruct a high-resolution iris image from LR counterparts. However, traditional dictionary based approaches such as K-Singular Value Decomposition (K-SVD) have two critical limitations: the assumption that the number of dictionary items has to be pre-defined explicitly; and the number of sparse coefficients has to be specified during the computational steps of Orthogonal Matching Pursuit (OMP). To overcome these limitations, Aljadaany *et al.* proposed to employ a modern non-parametric Bayesian approach, named Beta Process (BP), to build the discriminative over-complete dictionary and discover parameters automatically [146]. Experiments on a subset of the CASIA iris database shows improved recognition performance in comparison with using original LR images and linear interpolation.

3.3.2. Feature-domain super-resolution for iris

Similar to feature-domain super-resolution for face [49], Nguyen *et al.* also pointed out two problems with the existing super-resolution approaches for iris [51, 23, 52]:

- The aim of applying SR to biometrics is not for visual enhancement, but to improve recognition performance. Most existing SR approaches are designed to produce visual enhancement. *If recognition improvement is desired, why do we not focus on super-resolving only items essential for recognition?*

- Each biometric modality has its own characteristics. Most existing SR approaches for biometrics are general-scene SR approaches. *Can any specific information from biometric models be exploited to improve SR performance?*

From these observations, they proposed a Bayesian approach to estimate the HR coefficients from the LR coefficients [51]. The authors approach the problem of estimating HR coefficients by transforming the relationship between LR and HR images from the pixel domain to the PCA domain. By establishing the relationship between the HR coefficients and their LR counterparts, they can estimate the HR coefficients by a Bayes MAP probability estimator [51]. The reconstructed HR coefficients have been shown to improve the recognition performance of the iris biometrics system.

The major drawback of these approaches is that the linear features such as PCA and LDA are not optimal for recognition in comparison with the nonlinear Gabor-based features which have been shown to be one of the most discriminant features for face [147] and iris [130] at that time. The challenge of using these nonlinear features in super-resolution is the difficulty in formulating the relationship between the low-resolution features and the high-resolution features in the feature domain. Nguyen *et al.* [23, 52] proposed to break the encoding process into 2 steps: (1) computing the global response of the normalized image with a Gabor wavelet, (2) further encoding with

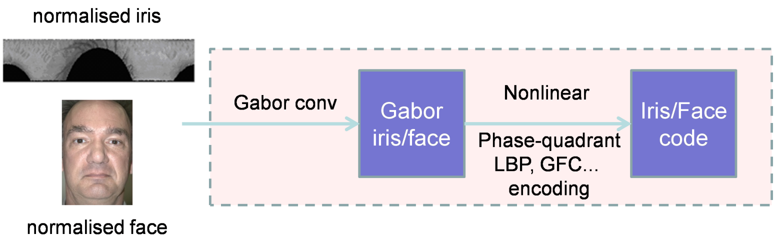
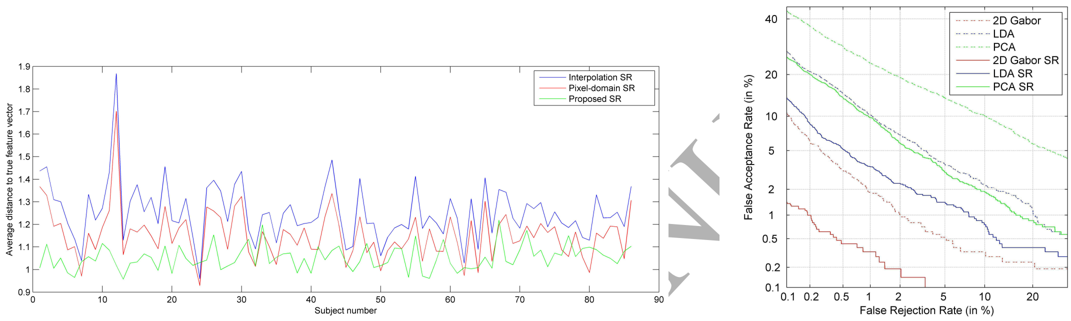


Figure 9: The common encoding flow in iris and face recognition systems using Gabor-based features. Both face and iris systems calculate the global response by convolving the whole image with the Gabor filter. After that, the Gabor images are further encoded with nonlinear steps such as Local Gabor Binary Pattern Histogram Sequence (LGBPHS). [23, 52]. Reprinted with permission from the authors.



(a) Distance to true vectors

(b) Linear vs. nonlinear features

Figure 10: One example of feature-domain super-resolution for iris [52]. The proposed feature-domain super-resolution using non-linear Gabor-based features has been shown to outperform its pixel-domain counterparts and other feature-domain super-resolution using linear features. Reprinted with permission from the authors.

other nonlinear techniques (e.g. phase-quadrant for iris [130], and LGBPHS for face [148]) as shown in Figure 9. Noting that the global Gabor response is linear, whilst the non-linearity of the overall encoding techniques results from the secondary encoding steps (e.g. phase-quadrant for iris, and LGBPHS for face), hence, they proposed to perform feature-domain SR on the global Gabor response (which they called intermediate features), rather than the final features to take advantage of the linear property of the global Gabor response. Since the intermediate features are linear, a simple MAP approach is applied to estimate the corresponding HR image. Improvements in the performance of iris recognition have been reported as illustrated in Figure 10.

Similarly, Alonso-Fernandez *et al.* proposed to change the PCA-feature-domain super-resolution process from using the whole image to using only local patches [149]. However, the major challenge of performing patch-based super-resolution is how to effectively preserve the geometry of the HR image when combining/concatenating HR patches to generate the final output. To resolve this challenge, later, they proposed using iterative neighbor embedding of local image patches to jointly consider the geometry of the LR and HR manifolds during the reconstruction process [150].

3.4. Super-resolution for gait

Human gait is a behavioral biometric modality which has seen successful applications for uncooperative human identification and/or human identification at a distance [151, 152]. It is based on body shape and the way a human walks. Gait recognition is one of the most interesting applications of super-resolution due to the nature of the gait matching process. While face and iris recognition, along with other biometrics, are based on matching two images (one test image and one train image), gait recognition, in contrast, is based on matching two video sequences (one test gait video sequence and one train gait video sequence). Consequently, in addition to the problem of low-resolution of the input (due to the low-resolution of the imaging device and the distance from the subject to the camera), gait recognition also faces challenges in matching a low frame-rate input video sequence to a normal/high-frame-rate gallery video sequence. Hence together with typical super-resolution approaches as discussed before, the gait research community also employ a concept named Temporal Super-Resolution.

3.4.1. Typical super-resolution approaches

In theory, all super-resolution approaches discussed in face (Section 3.1) and iris (Section 3.3) are applicable to super-resolve each frame of the gait video sequence to a higher resolution as a pre-processing step. Zhang *et al.* [153, 154] partitioned the gait frames into patches, then applying super-resolution with neighbor embedding to learn the HR counterpart of the LR test gait images, followed by a back-projection to make the LR and HR pairs more consistent. The SR approach employed is example-based. The neighbor embedding is used to preserve the neighborhood relationship between each point and its neighboring ones.

3.4.2. Temporal super-resolution approaches

While spatial super-resolution techniques enhance spatial resolution, temporal super-resolution techniques focus on improving temporal resolution in low frame-rate videos. Temporal super-resolution has been applied mainly for video editing and video compression. The concept "temporal super-resolution" was first introduced by Shechtman *et al.* [61] by proposing a space-time super-resolution approach to recover rapid dynamic events that occur faster than the regular frame-rate, and which are not visible (or else are observed incorrectly) in any of the input sequences, even if these are played in "slow-motion". In many applications such as sport videos, the regular frame-rate imaging devices could not reveal sufficient details of fast actions and are severely degraded by motion blur and motion aliasing. The authors proposed to perform joint spatial and temporal super-resolution by seeking a high-resolution space-time reconstruction video $\overrightarrow{h}$ from multiple low-resolution space-time measurements $\overrightarrow{l}$ from multiple cameras. This was achieved by minimizing the reconstruction error with directional space-time regularization terms,

$$min(||A\overrightarrow{h} - \overrightarrow{l}||^2 + ||W_x L_x \overrightarrow{h}||^2 + ||W_y L_y \overrightarrow{h}||^2 + ||W_t L_t \overrightarrow{h}||^2), \tag{33}$$

where $L_j (j = x, y, t)$ is a matrix capturing the second-order derivative operator in direction j and W_x is a diagonal weight matrix capturing the degree of the desired regularization at each space-time point in the direction j. The prominent property of this space-time SR is the trade-off between spatial and temporal resolution gains.

Temporal super-resolution can be split into two classes: reconstruction-based approaches and example-based approaches. Similar to the spatial domain, example-based approaches exploit the fact that small space-time patches of a natural video have the tendency to recur itself inside the

(a) Reconstructed temporal frames (b) EER

Figure 11: One example of temporal super-resolution for gait [158]. Reprinted with permission from the authors.

same video at multiple spatio-temporal scales [155] or by exploiting self-similar appearance of patches across different temporal resolution [156]. The temporal SR rates for the example-based approaches are quite limited (e.g. twice). Reconstruction-based approaches enhance temporal resolution by combining multiple videos of the same dynamic scene at sub-frame accuracy, either by simultaneously recording by multiple video cameras [157] or sequentially recording of multiple repetitions of a cyclic motion by a single camera [61].

In theory, all above approaches can be applied to low frame-rate gait videos. However, gait has its special characteristics and the main aim of applying temporal super-resolution for gait is also to improve recognition performance, hence a number of modifications can be added to make temporal super-resolution more effective. In [159, 160], Al-Huseiny *et al.* proposed a level-set approach to temporal interpolation to apply to gait recognition in low frame-rate videos. The authors extracted and represented boundary shapes by PCA decomposition. A set of coefficients is computed based on this PCA level set representation to quantify the contribution of each eigenvector to each shape,

$$\alpha_i = \gamma_k^T (u_i - \mu), \tag{34}$$

where γ_k is the matrix of the first k eigenvectors, u_i is the shape i^{th} and μ is the mean of all shapes. Since the coefficients, α_i, sequence are discrete due to frames. By fitting a cubic spline to the discrete sequence, the authors could infer the intermediate values for the missing shapes or to reconstruct new temporal views. The missing shapes or new temporal views can be reconstructed using k principal components weighted by a k-dimensional coefficient vector, $\hat{\alpha_i}$,

$$\hat{u_i} = \gamma_k \hat{\alpha_i} + \mu, \tag{35}$$

where α_i is the inferred coefficient. In a similar vein, Prismal *et al.* used linear interpolation of the Zernike moment values to infer new temporal views for temporal interpolation [161].

In [162], a periodic temporal SR method was employed to build a high frame-rate gait sequence period based on multiple periods of low frame-rate gait sequences. Based on [162], Akae *et al.* applied an exemplar of high frame-rate image sequences to improve the temporal SR quality [163]. A unified framework of example-based and reconstruction-based periodic temporal SR was proposed in [158, 164]. An energy function consists of four aspects: (1) data fitness between the interpolation and the input, (2) fitness between the control point matrix and the exemplar-based estimator, (3) smoothness of the periodic manifold in the silhouette eigen space, and (4) smoothness of the phase evolution based on the linear phase evolution prior. By minimizing the energy function, the control point matrix, coefficients of the Eigen gaits and phase sequence can be estimated to infer the high frame-rate output sequence. One example of the temporal super-resolved gait sequence and the recognition performance are presented in Figure 11.

3.5. Super-resolution for other modalities

Latent prints (fingerprint and palmprint), when captured at crime scenes, are normally of low quality and low-resolution. High-resolution images are required to increase the recognition accuracy [165, 166]. Despite a long history of latent prints, there is a limited number of published works on super-resolving latent-print images. From a traditional maximum likelihood super-resolution approach, Yuan *et al.* reconstructed a HR fingerprint image iteratively from multiple LR fingerprint images [167]. Observing that finding the optimal number of iterations greatly impacts the Signal-Noise-Ratio (SNR) of the super-resolved HR image, the authors proposed to employ early stopping regularization and a boundary constraint prior to make the SNR of the output HR image robust to the variation in the number of gradient descent steps. This helps to improve the accuracy of super-reconstructed HR images. However, the structure of the ridges and pores present in the fingerprint image, which have been a dominate feature in fingerprint recognition approaches [168], have not been efficiently considered to improve the quality of super-resolution reconstruction. Jiji *et al.* proposed a learning-based single-frame SR approach using contourlet transform [169]. The contourlet transform approach [170] has the capability to capture smoothness along contours coupled with learning the HR representation of an oriented edge primitive from the HR training set. Instead of focusing on the edge, Lian performed SR on the local structure of the image [171]. The local structure is estimated by a variational formulation method with a proposed adaptive filter reflecting both local pixel variance and global image information. Both this edge-based and structure-based SR approaches had been shown to improve the SNR of the reconstructed ridges in the super-resolved fingerprint image. Recently, Singh *et al.* and Bian *et al.* incorporated fingerprint structure information into a sparse-coding super-resolution approach [172, 173]. Rather than using one over-complete large dictionary of coupled HR and LR fingerprint patches, the authors proposed to cluster it into multiple sub-dictionaries. The sub-dictionaries are chosen based on the dominant orientation of the ridge structure within the patch.

Recognizing people by their ear has recently received significant attention in the literature [174, 175]. However, it is generally limited to indoor imaging conditions. When it comes to real-life applications such as surveillance, the ear images are usually of low quality and low-resolution, which paves the way for super-resolution. Luo *et al.* employed a sparse-coding super-resolution approach for ear images [176]. The dictionary is constructed by a Label Consistent K-SVD algorithm [177] to incorporate both reconstruction error and classification error to improve the recognition performance of the super-resolved ear images. This has been shown to improve the recognition performance in University of Notre Dame (UND) and University of Science and Technology Beijing (USTB) II Ear dataset [176].

4. Trends and Challenges

Super-resolution techniques have been employed successfully to improve the visual appearance of biometric samples and the recognition performance of biometric systems, especially for face, iris and gait modalities. Super-resolution, which attempts to reconstruct a high-resolution image from a number of low-resolution images, is a computationally complex and numerically ill-posed problem. The inverse solutions of super-resolution techniques are either based on learning the lost high frequency components from the training data or reconstructing them from multiple complementary low-resolution images. When there is only one low-resolution input image, the learning-based super-resolution is employed. In case of multiple low-resolution input images being available, the reconstruction-based super-resolution is usually applied. The super-resolution techniques can be performed in either the pixel domain or the frequency domain. In general, super-resolution is employed to improve the visual clarity of the images. However, in the case of biometrics, the major aim is to improve the recognition performance. Hence many approaches have been focused on performing super-resolution on the feature domain to reconstruct the high-resolution features. Performing super-resolution on the feature domain helps to not only reduce the effects of unnecessary details, which are eventually discarded later by the feature extraction process, but also incorporate specific priors from biometric models to limit the constraints and improve the recognition performance.

Super-resolution has been applied extensively to face, iris and gait recognition, but its application is still very limited in other biometric modalities. One of the long standing biometric modalities that has seen limited use of super-resolution are latent prints, including fingerprint and palmprint. This may be attributed to the contact imaging nature of these latent prints. However, in the presence of multiple instances of the same latent, e.g. multiple latent prints present at the crime scene, super-resolution techniques are useful to combine them to generate a fused, high-resolution image for improving the visual and recognition performance. Especially, with recent development of contactless fingerprint [178, 179] and palmprint [180, 181] approaches, super-resolution techniques developed in face, iris and gait can be utilized to improve the recognition performance. The major challenge of latent-print super-resolution is the non-uniform deformation of the captured images due to the curvature surface of the finger and the palm. Dealing with this non-uniform deformation requires encoding geometry information in the super-resolution process. Another modality that has seen limited use of super-resolution is ear biometrics. The ear is a promising candidate for super-resolution since it exhibits less variability with change in expression or orientation than the face [174]. Super-resolving ear images in both 2D or 3D would improve the operability and performance in real-life scenarios.

There are other non-traditional biometric modalities that have not seen super-resolution in use. For example, periocular, as a sub-region of the entire face, is also a potential candidate for applying super-resolution. Periocular is attracting more attention since it achieves a great trade-off between iris recognition and face recognition in terms of imaging range and recognition performance [182, 183, 184, 185]. Super-resolution approaches for the face can be migrated with little effort to the periocular region. Another promising candidate is soft-biometrics. Soft-biometrics relies on personal attributes such as gender, ethnicity, age, height, weight, eye color, scars, marks and tattoos to acquire some information about the identity of a person, but does not provide sufficient evidence to precisely determine the identity [186, 187]. As soft-biometrics are usually extracted and utilized in surveillance applications [188] where the evidence attributes exhibit the adverse influence of low-resolution issues due to the long distance, super-resolution techniques would be useful to improve both visual clarity and recognition performance. Similar to super-resolution for biometrics, super-resolution for soft-biometrics requires incorporating specific characteristics of the modality to achieve high performance on the reconstructed images/videos.

A major challenge applying super-resolution is the non-universality of the current approaches. While most proposed approaches have shown improvement in recognition performance, they are usually carefully tuned for a small group of datasets. Super-resolution techniques which perform well in diverse and complex forms of noise and variations in real-life low-resolution situations are yet to be developed. Most current approaches employ synthetic low-resolution data which are degraded from the high-resolution data with predefined degradation models. By degrading from the high-resolution data, super-resolution techniques can take advantage in terms of controlling the degradation factors with the predefined assumption made by the researchers. However, with regard to the diverse and complex nature of real-life applications, this degradation model may not reflect the real-life case. This raises the need for real-life and standard datasets rather than the artificial to assess and benchmark performance over a wide range of impairments. Advances in super-resolution with such data sets representing real-life Closed-Circuit Television (CCTV) camera feeds will be valuable for suspect identification through the use of biometrics for covert surveillance.

Recently, deep-learning-based super-resolution approaches have shown very promising reconstruction results [53, 108, 56]. A number of approaches have been introduced to face [113, 117, 115] and iris [139] but this area of research is still in its infancy since the specific prior in the face and iris structure has not been effectively incorporated. As deep learning techniques are extremely effective at feature learning from the data, the research community can capitalize on this capability to learn the specific priors of the biometric modalities without any pre-assumption from a human point of view. In addition, there are many different architectures of deep learning, including, e.g. CNNs, Recurrent Neural Networks, Reinforcement Learning, and Generative Adversarial Networks. Experimenting with these architectures with their advanced techniques to achieve optimal or near-optimal performance for deep learning based SR based systems is still an open and challenging question to the research community. In addition, most of the existing works still employ face super-resolution as a pre-processing step preceding the recognition step. Learning two steps jointly could also greatly improve the performance of both super-resolution and recognition.

5. Conclusions

Super-resolution is a core topic in computer vision, which has been deployed successfully for biometrics not only to improve the clarity and visual appearance of the images, but more recently and more importantly to improve the recognition performance of the system. Super-resolution enables the identification of subjects from significant stand-off distances which is valuable for both user-assisted authentication as well as covert suspect identification in surveillance applications. The combination of biometrics and super resolution also enables deeper analysis of the ever increasing visual data appearing in the world-wide web (such as in Youtube) and other public databases through identifying people from low-resolution imagery that is common in this data. There are specific advantages to evaluating super-resolution on biometrics, as the performance advantages can be precisely demonstrated. We have presented a comprehensive survey on the state-of-the-art super-resolution approaches for four major biometric modalities: face (2D+3D), iris, fingerprint and gait. We have approached the super-resolution problem from several different perspectives, including spatial domain and frequency domain, single and multiple input images, learning-based and reconstruction-based. Despite the success in these four major biometric modalities, applying super-resolution to biometrics still faces challenges in dealing with other long-standing modalities like latent print (e.g. fingerprint and palmprint) and emerging modalities (e.g. ear, periocular and soft-biometrics). In addition, applying super-resolution in real-life applications also requires better algorithms to deal with challenges in difficult and extreme imaging conditions, non-linear deformation of the subjects and the non-universality of the current approaches. This review, in addition to being useful to developers wanting to improve the performance of their biometric solutions, will serve as a catalyst to spur further research activities in this area, especially to address the remaining key challenges in the deployment of super-resolution for biometrics.

References

[1] M. Tistarelli, S. Li, R. Chellappa, Handbook of Remote Biometrics, Springer-Verlag London, UK, 2009.

[2] J. Neves, F. Narducci, S. Barra, H. Proença, Biometric recognition in surveillance scenarios: a survey, Artificial Intelligence Review 46 (2016) 515–541.

[3] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, A. Ross, Long range iris recognition: A survey, Pattern Recognition 72 (2017) 123–143.

[4] K. Nasrollahi, T. Moeslund, Super-resolution: A comprehensive survey, Machine Vision and Applications 25 (2014) 1423–1468.

[5] P. Milanfar, Super-resolution imaging, CRC Press, US, 2011.

[6] H. Zhang, L. Zhang, H. Shen, A super-resolution reconstruction algorithm for hyperspectral images, Signal Processing 92 (2012) 2082–2096.

[7] A. Boucher, P. Kyriakidis, C. Cronkite-Ratcliff, Geostatistical solutions for super-resolution land cover mapping, IEEE Transactions on Geoscience and Remote Sensing 46 (2008) 272–283.

[8] X. Zhu, R. Bamler, Demonstration of super-resolution for tomographic sar imaging in urban environment, IEEE Transactions on Geoscience and Remote Sensing 50 (2012) 3150–3157.

[9] P. Milanfar, Super-resolution imaging, CRC Press, 2010, Ch. New Applications of Super-Resolution in Medical Imaging, pp. 383–412.

[10] H. Greenspan, Super-resolution in medical imaging, Computer Journal 52 (2009) 43–63.

[11] Y. Huang, L. Shao, A. F. Frangi, Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[12] D. Capel, A. Zisserman, Automated mosaicing with super-resolution zoom, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 1998, pp. 885–891.

[13] P. Kramer, J. Benois-Pineau, J.-P. Domenger, Local object-based super-resolution mosaicing from low-resolution video, Signal Processing 91 (2011) 1771–80.

[14] B. Gunturk, Y. Altunbasak, R. Mersereau, Super-resolution reconstruction of compressed video using transform-domain statistics, IEEE Transactions on Image Processing 13 (2004) 33–43.

[15] K. Li, Y. Zhu, J. Yang, J. Jiang, Video super-resolution using an adaptive superpixel-guided autoregressive model, Pattern Recognition 51 (2016) 59–71.

[16] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, W. Shi, Real-time video super-resolution with spatio-temporal networks and motion compensation, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[17] K. Nasrollahi, S. Escalera, P. Rasti, G. Anbarjafari, X. Baro, H. J. Escalante, T. B. Moeslund, Deep learning based super-resolution for improved action recognition, in: 2015 International Conference on Image Processing Theory, Tools and Applications (IPTA), 2015, pp. 67–72.

[18] M. S. Ryoo, B. Rothrock, C. Fleming, Privacy-preserving egocentric activity recognition from extreme low resolution, CoRR abs/1604.03196.

[19] C. Hong, J. Yu, J. Wan, D. Tao, M. Wang, Multimodal deep autoencoder for human pose recovery, IEEE Transactions on Image Processing 24 (12) (2015) 5659–5670.

[20] C. Hong, J. Yu, D. Tao, M. Wang, Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval, IEEE Transactions on Industrial Electronics 62 (6) (2015) 3742–3751.

[21] S. Baker, T. Kanade, Limits on super-resolution and how to break them, IEEE Transaction on Pattern Analysis and Machine Intelligence 24 (2002) 1167–1183.

[22] N. Akae, Y. Makihara, Y. Yagi, Gait recognition using periodic temporal super resolution for low frame-rate videos, in: International Joint Conference on Biometrics (IJCB), 2011, pp. 1–7.

[23] K. Nguyen, S. Sridharan, S. Denman, C. Fookes, Feature-domain super-resolution framework for gabor-based face and iris recognition, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), US, 2012, pp. 2642–2649.

[24] A. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Transactions on Circuits and Systems for Video Technology 14 (2004) 4–20.

[25] A. Jain, P. Flynn, A. Ross, Handbook of Biometrics, Springer-Verlag New York, Inc., US, 2007.

[26] S. Park, M. Park, M. Kang, Super-resolution image reconstruction: a technical overview, IEEE Signal Processing Magazine 20 (2003) 21–36.

[27] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, L. Zhang, Image super-resolution: The techniques, applications and future, Signal Processing 128 (2016) 389–408.

[28] N. Wang, D. Tao, X. Gao, X. Li, J. Li, A comprehensive survey to face hallucination, International Journal of Computer Vision (IJCV) 106 (2014) 9–30.

[29] R. Gerchberg, Super-resolution through error energy reduction, Optica Acta: International Journal of Optics 21 (1974) 709–720.

[30] P. Chatterjee, S. Mukherjee, S. Chaudhuri, G. Seetharaman, Application of papoulis-gerchberg method in image super-resolution and inpainting, Computer Journal 52 (2009) 80–89.

[31] R. Hardie, K. Barnard, E. Armstrong, Joint map registration and high-resolution image estimation using a sequence of undersampled images, IEEE Transactions on Image Processing 6 (1997) 1621–33.

[32] A. Panagiotopoulou, V. Anastassopoulos, Regularized super-resolution image reconstruction employing robust error norms, Optical Engineering 48 (2009) 117004–117018.

[33] A. Patti, Y. Altunbasak, Artifact reduction for set theoretic super resolution image reconstruction with edge adaptive constraints and higher-order interpolants, IEEE Transactions on Image Processing 10 (2001) 179–86.

[34] M. Elad, A. Feuer, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images, IEEE Transactions on Image Processing 6 (1997) 1646–58.

[35] M. Irani, S. Peleg, Improving resolution by image registration, CVGIP: Graphical Models and Image Processing 53 (1991) 231–9.

[36] G. Gilboa, N. Sochen, Y. Zeevi, Forward-and-backward diffusion processes for adaptive image enhancement and denoising, IEEE Transactions on Image Processing 11 (2002) 689–703.

[37] R. Tsai, T. Huang, Multi-frame image restoration and registration, Advances in Computer Vision and Image Processing (1984) 317–339.

[38] H. Demirel, G. Anbarjafari, Image resolution enhancement by using discrete and stationary wavelet decomposition, IEEE Transactions on Image Processing 20 (2011) 1458–1460.

[39] P. Sen, S. Darabi, Compressive image super-resolution, in: 43rd Asilomar Conference on Signals, Systems and Computers, 2009, pp. 1235–1242.

[40] G. Fahmy, Super-resolution construction of iris images from a visual low resolution face video, in: International Symposium on Signal Processing and Its Applications (ISSPA), 2007, pp. 1–4.

[41] K. Nguyen, C. Fookes, S. Sridharan, S. Denman, Focus-score weighted super-resolution for uncooperative iris recognition at a distance and on the move, in: 25th International Conference of Image and Vision Computing New Zealand (IVCNZ), 2010, pp. 1–8.

[42] K. Nguyen, C. Fookes, S. Sridharan, S. Denman, Quality-driven super-resolution for less constrained iris recognition at a distance and on the move, IEEE Transactions on Information Forensics and Security 6 (2011) 1248–1258.

[43] R. Barnard, V. Pauca, T. Torgersen, R. Plemmons, S. Prasad, J. van der Gracht, J. Nagy, J. Chung, G. Behrmann, S. Mathews, M. Mirotznik, High-resolution iris image reconstruction from low-resolution imagery, in: SPIE Advanced Signal Processing Algorithms, Architectures, and Implementations XVI, Vol. 6313, US, 2006, pp. 63130–1.

[44] J. Huang, L. Ma, T. Tan, Y. Wang, Learning based resolution enhancement of iris images, in: British Machine Vision Conference (BMVC), Vol. 1, US, 2003, pp. 1–10.

[45] K. Shin, K. Park, B. Kang, S. Park, Super-resolution method based on multiple multi-layer perceptrons for iris recognition, in: International Conference on Ubiquitous Information Technologies Applications, 2009, pp. 1–5.

[46] S. Li, A. Jain, Encyclopedia of Biometrics, Springer US, 2009, Ch. Iris super-resolution, pp. 856–859.

[47] J. Zhang, L. Zhang, L. Xiang, Y. Shao, G. Wu, X. Zhou, D. Shen, Q. Wang, Brain atlas fusion from high-thickness diagnostic magnetic resonance images by learning-based super-resolution, Pattern Recognition 63 (2017) 531–541.

[48] R. He, Z. Zhang, Locally affine patch mapping and global refinement for image super-resolution, Pattern Recognition 44 (9) (2011) 2210–2219.

[49] B. Gunturk, A. Batur, Y. Altunbasak, M. Hayes, R. Mersereau, Eigenface-domain super-resolution for face recognition, IEEE Transactions on Image Processing 12 (2003) 597–606.

[50] K. Jia, S. Gong, Generalized face super-resolution, IEEE Transactions on Image Processing 17 (2008) 873–886.

[51] K. Nguyen, C. Fookes, S. Sridharan, S. Denman, Feature-domain super-resolution for iris recognition, in: IEEE International Conference on Image Processing (ICIP), BE, 2011, pp. 3258–3261.

[52] K. Nguyen, C. Fookes, S. Sridharan, S. Denman, Feature-domain super-resolution for iris recognition, Computer Vision and Image Understanding 117 (2013) 1526–1535.

[53] C. Dong, C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision (ECCV), Vol. IV, CH, 2014, pp. 184–99.

[54] Z. Cui, H. Chang, S. Shan, B. Zhong, X. Chen, Deep network cascade for image super-resolution, in: European Conference on Computer Vision (ECCV), CH, 2014, pp. 49–64.

[55] W. Shi, J. Caballero, F. Huszar, J. Totz, A. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 11874–1883.

[56] J. Kim, J. Lee, K. Lee, Accurate image super-resolution using very deep convolutional networks, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1646–1654.

[57] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-realistic single image super-resolution using a generative adversarial network, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[58] L. Wang, Z. Huang, Y. Gong, C. Pan, Ensemble based deep networks for image super-resolution, Pattern Recognition 68 (2017) 191–198.

[59] E. Mjolsness, Neural networks, pattern recognition, and fingerprint hallucination, Ph.D. thesis, California Institute of Technology, US (1986).

[60] S. Baker, T. Kanade, Hallucinating faces, in: IEEE International Conference on Automatic Face and Gesture Recognition (FG), 2000, pp. 83–88.

[61] E. Shechtman, Y. Caspi, M. Irani, Space-time super-resolution, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (2005) 531–545.

[62] A. Lemieux, M. Parizeau, Experiments on eigenfaces robustness, in: International Conference on Pattern Recognition (ICPR), Vol. 1, 2002, pp. 421–424.

[63] J. Wang, C. Zhang, H. Shum, Face image resolution versus face recognition performance based on two global methods, in: Asian Conference on Computer Vision (ACCV), 2004.

[64] B. Boom, G. Beumer, L. Spreeuwers, R. Veldhuis, The effect of image resolution on the performance of a face recognition system, in: 9th International Conference on Control, Automation, Robotics and Vision (ICARCV), 2006.

[65] F. Lin, Super-resolution image processing with application to face recognition, Ph.D. thesis, Queensland University of Technology, AU (2008).

[66] C. Fookes, F. Lin, V. Chandran, S. Sridharan, Evaluation of image resolution and super-resolution on face recognition performance, Journal of Visual Communication and Image Representation 23 (2012) 75–93.

[67] Y. Lui, D. Bolme, B. Draper, J. Beveridge, G. Givens, P. Phillips, A meta-analysis of face recognition covariates, in: IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS), US, 2009, pp. 1–8.

[68] P. Hennings-Yeomans, S. Baker, B. Kumar, Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.

[69] K. Jia, S. Gong, Multi-modal tensor face for simultaneous super-resolution and recognition, in: IEEE International Conference on Computer Vision (ICCV), Vol. 2, 2005, pp. 1683–1690.

[70] C. Liu, H. Shum, C. Zhang, A two-step approach to hallucinating faces: global parametric model and local nonparametric model, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, 2001.

[71] W. Zou, P. Yuen, Very low resolution face recognition problem, IEEE Transactions on Image Processing 21 (2012) 327–40.

[72] C. Liu, H. Shum, W. Freeman, Face hallucination: Theory and practice, International Journal of Computer Vision 75 (2007) 115–134.

[73] W. Zhang, W. Cham, Learning-based face hallucination in dct domain, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.

[74] W. Zhang, W. Cham, Hallucinating face in the dct domain, IEEE Transactions on Image Processing 20 (2011) 2769–2779.

[75] J. Park, S. Lee, An example-based face hallucination method for single-frame, low-resolution facial images, IEEE Transactions on Image Processing 17 (2008) 1806–1816.

[76] X. Ma, J. Zhang, C. Qi, An example-based two-step face hallucination method through coefficient learning, in: Lecture Notes in Computer Science, Vol. 5627, CA, 2009, pp. 471–480.

[77] W. Freeman, E. Pasztor, Learning low-level vision, in: IEEE International Conference on Computer Vision (ICCV), Vol. 2, 1999, pp. 1182–1189.

[78] W. Freeman, T. Jones, E. Pasztor, Example-based super-resolution, IEEE Computer Graphics and Applications 22 (2002) 56–65.

[79] T. Stephenson, T. Chen, Adaptive markov random fields for example-based super-resolution of faces, EURASIP Journal on Advances in Signal Processing 2006 (2006) 1–11.

[80] A. Akyol, M. Gokmen, Super-resolution reconstruction of faces by enhanced global models of shape and texture, Pattern Recognition 45 (12) (2012) 4103–4116.

[81] J. Matthews, S. Baker, Active appearance models revisited, International Journal of Computer Vision 60 (2) (2004) 135–164.

[82] X. Wang, X. Tang, Hallucinating face by eigentransformation, IEEE Transactions on Systems, Man, and Cybernetics: Part C 35 (2005) 425–34.

[83] H. Huang, H. He, X. Fan, J. Zhang, Super-resolution of human face image using canonical correlation analysis, Pattern Recognition 43 (7) (2010) 2532–2543.

[84] W. Liu, D. Lin, X. Tang, Hallucinating faces: Tensorpatch super-resolution and coupled residue compensation, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, 2005, pp. 478–484.

[85] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2009) 210–227.

[86] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution as sparse representation of raw image patches, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2008, pp. 1–8.

[87] J. Yang, J. Wright, T. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Transactions on Image Processing 19 (2010) 2861–2873.

[88] J. Yang, H. Tang, Y. Ma, T. Huang, Face hallucination via sparse coding, in: IEEE International Conference on Image Processing (ICIP), 2008, pp. 1264–1267.

[89] G. Gao, J. Yang, Sparse representation based face image super-resolution, in: International Conference on Intelligent Science and Intelligent Data Engineering, Vol. 7202, 2012, pp. 303–308.

[90] G. Gao, J. Yang, A novel sparse representation based framework for face image super-resolution, Neurocomputing 134 (2014) 92–99.

[91] Z. Hu, W. Li, Q. Tang, Y. Chen, Face super resolution by patch-based sparse coding, in: International Conference on Computer Science Service System (CSSS), 2012, pp. 1522–1525.

[92] E. Bilgazyev, B. Efraty, S. Shah, T. Kakadiaris, Sparse representation-based super resolution for face recognition at a distance, in: British Machine Vision Conference (BMVC), 2011, pp. 52.1–52.11.

[93] Y. Li, C. Cai, G. Qiu, K. Lam, Face hallucination based on sparse local-pixel structure, Pattern Recognition 47 (2014) 1261–1270.

[94] Z. Jia, H. Wang, Z. Xiong, A. Finn, Fast face hallucination with sparse representation for video surveillance, in: Asian Conference on Pattern Recognition (ACPR), 2011, pp. 179–183.

[95] J. Jiang, C. Chen, J. Ma, Z. Wang, Z. Wang, R. Hu, Srlsp: A face image super-resolution algorithm using smooth regression with local structure prior, IEEE Transactions on Multimedia 19 (2017) 27–40.

[96] C. Jung, L. Jiao, B. Liu, M. Gong, Position-patch based face hallucination using convex optimization, IEEE Signal Processing Letters 18 (2011) 367–370.

[97] J. Jiang, R. Hu, Z. Wang, Z. Xiong, Z. Han, Support-driven sparse coding for face hallucination, in: IEEE International Symposium on Circuits and Systems (ISCAS), 2013, pp. 2980–2983.

[98] J. Jiang, R. Hu, Z. Han, Z. Wang, J. Chen, Two-step superresolution approach for surveillance face image through radial basis function-partial least squares regression and locality-induced sparse representation, Journal of Electronic Imaging 22 (2013) 041120.

[99] K. Huang, R. Hu, Z. Han, T. Lu, J. Jiang, F. Wang, Face image superresolution via locality preserving projection and sparse coding, Journal of Software 8.

[100] K. Huang, R. Hu, Z. Han, T. Lu, J. Jiang, F. Wang, Graph regularized sparse coding for face hallucination, Information Technology Journal 13 (2014) 1883–1887.

[101] C. Jung, L. Jiao, B. Liu, M. Gong, Position-patch based face hallucination using convex optimization, IEEE Signal Processing Letters 18 (2011) 367–370.

[102] F. Juefei-Xu, M. Savvides, Single face image super-resolution via solo dictionary learning, in: IEEE International Conference on Image Processing (ICIP), 2015, pp. 2239–2243.

[103] L. Yann, B. Yoshua, H. Geoffrey, Deep learning, Nature 521 (2015) 436444.

[104] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks 61 (2015) 85–117.

[105] Y. Bengio, A. Courville, P. Vincent, Representation learning: A review and new perspectives, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (8) (2013) 1798–1828.

[106] C. Dong, C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (2016) 295–307.

[107] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, in: IEEE International Conference on Computer Vision (ICCV), 2015, pp. 370–378.

[108] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, T. Huang, Robust single image super-resolution via deep networks with sparse prior, IEEE Transactions on Image Processing 25 (2016) 3194–3207.

[109] J. Kim, J. Lee, K. Lee, Deeply-recursive convolutional network for image super-resolution, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[110] Y. Tai, J. Yang, X. Liu, Image super-resolution via deep recursive residual network, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[111] W.-S. Lai, J.-B. Huang, N. Ahuja, M.-H. Yang, Deep laplacian pyramid networks for fast and accurate super-resolution, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[112] C. Dong, C. Loy, X. Tang, Accelerating the super-resolution convolutional neural network, in: European Conference on Computer Vision (ECCV), 2016.

[113] P. Rasti, T. Uiboupin, S. Escalera, G. Anbarjafari, Convolutional neural network super resolution for face recognition in surveillance monitoring, in: 9th International Conference on Articulated Motion and Deformable Objects (AMDO), 2016, pp. 175–184.

[114] G. Berger, C. Peyrard, M. Baccouche, Boosting face recognition via neural super-resolution, in: 24th European Symposium on Artificial Neural Networks, BE, 2016, pp. 399–404.

[115] W. Ko, S. Chien, Patch-based face hallucination with multitask deep neural network, in: IEEE International Conference on Multimedia and Expo (ICME), US, 2016, pp. 1–6.

[116] E. Zhou, H. Fan, Z. Cao, Y. Jiang, Q. Yin, Learning face hallucination in the wild, in: 29th AAAI Conference on Artificial Intelligence, 2015, pp. 3871–3877.

[117] S. Zhu, S. Liu, C. Loy, X. Tang, Deep cascaded bi-network for face hallucination, in: European Conference on Computer Vision (ECCV), 2016, pp. 614–630.

[118] Q. Cao, L. Lin, Y. Shi, X. Liang, G. Li, Attention-aware face hallucination via deep reinforcement learning, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[119] O. Tuzel, Y. Taguchi, J. R. Hershey, Global-local face upsampling network, CoRR abs/1603.07235.

[120] S. Peng, G. Pan, Z. Wu, Learning-based super-resolution of 3d face model, in: IEEE International Conference on Image Processing (ICIP), Vol. 2, IT, 2005, pp. 382–385.

[121] G. Pan, S. Han, Z. Wu, Y. Wang, Super-resolution of 3d face, in: European Conference on Computer Vision (ECCV), DE, 2006, pp. 389–401.

[122] S. Berretti, A. Del Bimbo, P. Pala, Superfaces: A super-resolution model for 3d faces, in: European Conference on Computer Vision (ECCV), Vol. I, DE, 2012, pp. 73–82.

[123] S. Berretti, P. Pala, A. del Bimbo, Face recognition by super-resolved 3d models from consumer depth cameras, IEEE Transactions on Information Forensics and Security 9 (2014) 1436–49.

[124] E. Bondi, P. Pala, S. Berretti, A. D. Bimbo, Reconstructing high-resolution face models from kinect depth sequences acquired in uncooperative contexts, in: IEEE International Conference on Automatic Face and Gesture Recognition (FG), Vol. 07, 2015, pp. 1–6.

[125] K. Ouji, M. Ardabilian, L. Chen, F. Ghorbel, 3d deformable super-resolution for multi-camera 3d face scanning, Journal of Mathematical Imaging and Vision 47 (2013) 124–37.

[126] J. Yu, B. Bhanu, Y. Xu, A. Roy-Chowdhury, Super-resolved facial texture under changing pose and illumination, in: IEEE International Conference on Image Processing (ICIP), Vol. 3, 2007, pp. 553–556.

[127] P. Mortazavian, J. Kittler, W. Christmas, A 3-d assisted generative model for facial texture super-resolution, in: IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS), 2009, pp. 1–7.

[128] P. Mortazavian, J. Kittler, W. Christmas, 3d-assisted facial texture super-resolution, in: British Machine Vision Conference (BMVC), UK, 2009.

[129] V. Blanz, T. Vetter, A morphable model for the synthesis of 3d faces, in: 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), US, 1999, pp. 187–194.

[130] J. Daugman, New methods in iris recognition, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 37 (2007) 1167–1175.

[131] NIST, Irex iv - evaluation of iris identification algorithms, Tech. Rep. NIST Interagency Report 7949, National Institute of Science and Technology, US (2013).

[132] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, W. Zhao, Iris on the move: acquisition of images for iris recognition in less constrained environments, Proceedings of the IEEE 94 (2006) 1936–47.

[133] F. Bashir, P. Casaverde, D. Usher, M. Friedman, Eagle-eyes: A system for iris recognition at a distance, IEEE International Conference on Technologies for Homeland Security (HST) (2008) 426–431.

[134] F. Bashir, D. Usher, P. Casaverde, M. Friedman, Video surveillance for biometrics: long-range multi-biometric system, IEEE Fifth International Conference on Advanced Video and Signal-based Surveillance (AVSS) (2008) 175–82.

[135] S. Venugopalan, U. Prasad, K. Harun, K. Neblett, D. Toomey, J. Heyman, M. Savvides, Long range iris acquisition system for stationary and mobile subjects, in: International Joint Conference on Biometrics (IJCB), 2011, pp. 1–8.

[136] NIST, Biometric data interchange formats, part 6: Iris image data, Tech. Rep. ISO/IEC 19794-6:2005, National Institute of Science and Technology, US (2005).

[137] N. Kalka, J. Zuo, N. Schmid, B. Cukic, Estimating and fusing quality factors for iris biometric images, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 40 (2010) 509–524.

[138] S. Schulter, C. Leistner, H. Bischof, Fast and accurate image upscaling with super-resolution forests, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3791–3799.

[139] Q. Zhang, H. Li, Z. He, Z. Sun, Image super-resolution for mobile iris recognition, in: Chinese Conference on Biometric Recognition (CCBR), 2016, pp. 399–406.

[140] A. Deshpande, P. P. Patavardhan, Super resolution and recognition of long range captured multi-frame iris images, IET Biometrics 6 (5) (2017) 360–368.

[141] A. Deshpande, P. Patavardhan, Multi-frame super-resolution for long range captured iris polar image, IET Biometrics 6 (2) (2017) 108–116.

[142] B. Zitov, J. Flusser, Image registration methods: a survey, Image and Vision Computing 21 (2003) 977–1000.

[143] N. Othman, N. Houmani, B. Dorizzi, Improving video-based iris recognition via local quality weighted super resolution, in: International Conference on Pattern Recognition Applications and Methods (ICPRAM), ES, 2013, pp. 623–629.

[144] N. Othman, B. Dorizzi, Impact of quality-based fusion techniques for video-based iris recognition at a distance, IEEE Transactions on Information Forensics and Security 10 (2015) 1590–1602.

[145] S. Hsieh, Y. Li, C. Tien, C. Chang, Extending the capture volume of an iris recognition system using wavefront coding and super-resolution, IEEE Transactions on Cybernetics 46 (2016) 3342–3350.

[146] R. Aljadaany, K. Luu, S. Venugopalan, M. Savvides, Iris super-resolution via nonparametric over-complete dictionary learning, in: IEEE International Conference on Image Processing (ICIP), 2015, pp. 3856–3860.

[147] A. Serrano, I. Diego, C. Conde, E. Cabello, Recent advances in face biometrics with gabor wavelets: A review, Pattern Recognition Letters 31 (2010) 372–81.

[148] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition, in: IEEE International Conference on Computer Vision (ICCV), Vol. 1, 2005, pp. 786–791.

[149] F. Alonso-Fernandez, R. Farrugia, J. Bigun, Very low-resolution iris recognition via eigen-patch super-resolution and matcher fusion, in: IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS), 2016, pp. 1–8.

[150] F. Alonso-Fernandez, R. A. Farrugia, J. Bigun, Iris super-resolution using iterative neighbor embedding, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 655–663.

[151] P. Larsen, E. Simonsen, N. Lynnerup, Gait analysis in forensic medicine, Journal of Forensic Sciences 6491 (2007) 64910–1.

[152] T. Lee, M. Belkhatir, S. Sanei, A comprehensive review of past and present vision-based techniques for gait recognition, Multimedia Tools and Applications 72 (2014) 2833–2869.

[153] J. Zhang, Y. Cheng, C. Chen, Low resolution gait recognition with high frequency super resolution, in: 10th Pacific Rim International Conference on Artificial Intelligence Trends (PRICAI), DE, 2008, pp. 533–43.

[154] J. Zhang, J. Pu, C. Chen, R. Fleischer, Low-resolution gait recognition, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 40 (2010) 986–996.

[155] O. Shahar, A. Faktor, M. Irani, Space-time super-resolution from a single video, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), US, 2011, pp. 3353–60.

[156] M. Shimano, T. Okabe, I. Sato, Y. Sato, Video temporal super-resolution based on self-similarity, in: Asian Conference on Computer Vision (ACCV), Vol. 1, DE, 2011, pp. 93–106.

[157] A. Agrawal, M. Gupta, A. Veeraraghavan, S. Narasimhan, Optimal coded sampling for temporal super-resolution, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), US, 2010, pp. 599–606.

[158] N. Akae, A. Mansur, Y. Makihara, Y. Yagi, Video from nearly still: An application to low frame-rate gait recognition, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1537–1543.

[159] M. Al-Huseiny, S. Mahmoodi, M. Nixon, Level set gait analysis for synthesis and reconstruction, in: 5th International Symposium in Visual Computing (ISVC), Vol. 2, DE, 2009, pp. 377–86.

[160] M. Al-Huseiny, S. Mahmoodi, M. Nixon, Gait learning-based regenerative model: A level set approach, in: International Conference on Pattern Recognition (ICPR), 2010, pp. 2644–2647.

[161] S. Prismall, M. Nixon, J. Carter, Novel temporal views of moving objects for gait biometrics, in: 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), DE, 2003, pp. 725–33.

[162] Y. Makihara, A. Mori, Y. Yagi, Temporal super resolution from a single quasi-periodic image sequence based on phase registration, in: Asian Conference on Computer Vision (ACCV), Vol. 1, DE, 2011, pp. 107–20.

[163] N. Akae, Y. Makihara, Y. Yagi, Gait recognition using periodic temporal super resolution for low frame-rate videos, in: International Joint Conference on Biometrics (IJCB), US, 2011, pp. 1–7.

[164] Y. Makihara, Towards robust gait recognition, in: IAPR Asian Conference on Pattern Recognition (ACPR), 2013, pp. 18–22.

[165] A. Jain, Y. Chen, M. Demirkus, Pores and ridges: High-resolution fingerprint matching using level 3 features, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2007) 15–27.

[166] Q. Zhao, D. Zhang, L. Zhang, N. Luo, High resolution partial fingerprint alignment using porevalley descriptors, Pattern Recognition (PR) 43 (2010) 1050–1061.

[167] Z. Yuan, J. Wu, S. Kamata, A. Ahrary, P. Yan, Fingerprint image enhancement by super resolution with early stopping, in: IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS), Vol. 4, 2009, pp. 527–531.

[168] D. Mario, D. Maltoni, Fingerprint Analysis and Representation, Springer London, London, 2009, pp. 97–166.

[169] C. Jiji, S. Chaudhuri, Single-frame image super-resolution through contourlet learning, EURASIP Journal on Advances in Signal Processing (2006) 235–235.

[170] M. Do, M. Vetterli, The contourlet transform: An efficient directional multiresolution image representation, IEEE Transactions on Image Processing 14 (2005) 2091–2106.

[171] H. Lian, Variational local structure estimation for image super-resolution, in: IEEE International Conference on Image Processing (ICIP), 2006, pp. 1721–1724.

[172] K. Singh, A. Gupta, R. Kapoor, Fingerprint image super-resolution via ridge orientation-based clustered coupled sparse dictionaries, Journal of Electronic Imaging 24 (2015) 043015.

[173] W. Bian, S. Ding, Y. Xue, Fingerprint image super resolution using sparse representation with ridge pattern prior by classification coupled dictionaries, IET Biometrics.

[174] A. Abaza, A. Ross, C. Hebert, M. Harrison, M. Nixon, A survey on ear biometrics, ACM Computing Survey 45 (2013) 22:1–22:35.

[175] A. Pflug, C. Busch, Ear biometrics: a survey of detection, feature extraction and recognition methods, IET Biometrics 1 (2012) 114–129.

[176] S. Luo, Z. Mu, B. Zhang, Discriminative Super-Resolution Method for Low-Resolution Ear Recognition, Springer International Publishing, 2014, pp. 442–450.

[177] Z. Jiang, Z. Lin, L. Davis, Label consistent k-svd: Learning a discriminative dictionary for recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (2013) 2651–2664.

[178] G. Abramovich, K. Harding, S. Manickam, J. Czechowski, V. Paruchuru, R. Tait, C. Nafis, A. Vemury, Mobile, contactless, single-shot, fingerprint capture system, in: SPIE Biometric Technology for Human Identification VII, Vol. 7667, US, 2010.

[179] A. Kumar, C. Kwong, Towards contactless, low-cost and accurate 3d fingerprint identification, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2013, pp. 3438–3443.

[180] K. Ito, T. Sato, S. Aoyama, S. Sakai, S. Yusa, T. Aoki, Palm region extraction for contactless palmprint recognition, in: International Conference on Biometrics (ICB), US, 2015, pp. 334–40.

[181] A. Morales, M. Ferrer, A. Kumar, Towards contactless palmprint authentication, IET Computer Vision 5 (2011) 407–16.

[182] U. Park, R. R. Jillela, A. Ross, A. K. Jain, Periocular biometrics in the visible spectrum, IEEE Transactions on Information Forensics and Security 6 (2011) 96–106.

[183] F. M. Algashaam, K. Nguyen, V. Chandran, J. Banks, Elliptical higher-order-spectra periocular code, IEEE Access 5 (2017) 6978–6988.

[184] F. M. Algashaam, K. Nguyen, M. Alkanhal, V. Chandran, W. Boles, J. Banks, Multispectral periocular classification with multimodal compact multi-linear pooling, IEEE Access 5 (2017) 14572–14578.

[185] Z. Zhao, A. Kumar, Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network, IEEE Transactions on Information Forensics and Security 12 (2017) 1017–1030.

[186] A. Dantcheva, P. Elia, A. Ross, What else does your biometric data reveal? a survey on soft biometrics, IEEE Transactions on Information Forensics and Security 11 (2016) 441–467.

[187] M. Nixon, P. Correia, K. Nasrollahi, T. Moeslund, A. Hadid, M. Tistarelli, On soft biometrics, Pattern Recognition Letters (PRL) 68 (2015) 218–230.

[188] D. Reid, S. Samangooei, C. Chen, M. Nixon, A. Ross, Soft biometrics for surveillance: An overview, Handbook of Statistics 31 (2013) 327–352.

Author Biography

Kien Nguyen is a Postdoctoral Research Fellow at Queensland University of Technology (QUT, Australia). He has been researching biometrics for 7 years, including 4 years of his PhD at QUT. His research interests are applications of computer vision and deep learning techniques for biometrics and scene understanding.

Clinton Fookes is a Professor in Vision & Signal Processing at the Queensland University of Technology. He holds a BEng (Aerospace/Avionics), an MBA, and a PhD in computer vision. He is a Senior Member of the IEEE, an AIPS Young Tall Poppy, an Australian Museum Eureka Prize winner, and a Senior Fulbright Scholar.

Sridha Sridharan obtained his MSc degree from the University of Manchester, UK and his PhD degree from University of New South Wales, Australia. He is currently a Professor at Queensland University of Technology (QUT) where he leads the research program in Speech, Audio, Image and Video Technologies (SAIVT).

Massimo Tistarelli is a Professor and Director of the Computer Vision Laboratory at the University of Sassari, Italy. He is a world-recognized leading researcher in biometrics, leading several National and European projects on biometrics. He is the Principal Editor of the Springer book "Handbook of Remote Biometrics".

Mark Nixon is a Professor in University of Southampton, UK. He was Principal Investigator on the DARPA supported project Automatic Gait Recognition and the EU-funded Tabula Rasa project. He is co-authors the book "Human ID based on Gait". Mark is a member of IAPR TC4 Biometrics and IEEE Biometrics Council.