Face biometrics using self-resolution approach

A COURSE PROJECT REPORT

By

Sri Krishna Siddhardha Varma K. (RA2011030010003)
Medhavi Mehta (RA2011030010004)
Harshitha Devi Gokaraju (RA2011030010020)
Devshree Moghe (RA2011030010049)

Under the guidance of

Ms.M.Vijaya Lakshmi

In partial fulfilment for the Course

of

18CSE357 - BIOMETRICS

in Networking and Communications

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Kattankulathur, Chenpalpattu District

NOVEMBER 2022

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this mini project report titled "**Face biometrics using self-resolution approach"** is the bonafide work done by Sri Krishna Siddhardha Varma K. (RA2011030010003), Medhavi Mehta (RA2011030010004), Harshitha Devi Gokaraju (RA2011030010020) and Devshree Moghe (RA2011030010049) who carried out the mini project work and Laboratory exercises under my supervision for 18CSE357 - BIOMETRICS

Certified further, that to the best of my knowledge the work reported herein does not form part of any other work.

Ms. M. Vijaya Lakshmi
ASSISTANT PROFESSOR
Department of Computing Technologies

ACKNOWLEDGMENT

We express our heartfelt thanks to our honorable **Vice Chancellor Dr. C. MUTHAMIZHCHELVAN**, for being the beacon in all our endeavors.

We would like to express my warmth of gratitude to our **Registrar Dr. S. Ponnusamy,** for his encouragement.

We express our profound gratitude to our **Dean (College of Engineering and Technology) Dr. T. V.Gopal,** for bringing out novelty in all executions.

We would like to express my heartfelt thanks to Chairperson, School of Computing **Dr. Revathi Venkataraman,** for imparting confidence to complete our course project.

We wish to express my sincere thanks to **Course Audit Professor Dr.Annapurani Panaiyappan, Professor and Head, Department of Networking and Communications** and **Course Coordinators** for their constant encouragement and support.

We are highly thankful to our Course project Faculty Ms.M.Vijaya Lakshmi, Assistant Professor, Department of Computing Technologies for her assistance, timely suggestion and guidance throughout the duration of this course project.

Finally, we thank our parents and friends near and dear ones who directly and indirectly contributed to the successful completion of our project. Above all, I thank the almighty for showering his blessings on me to complete my Course project.

TABLE OF CONTENTS

Abstract

Selfie-based biometrics has great potential for a wide range of applications since, e.g. periocular verification is contactless and is safe to use in pandemics such as COVID-19, when a major portion of a face is covered by a facial mask. Despite its advantages, selfie-based biometrics presents challenges since there is limited control over data acquisition at different distances. Therefore, Super-Resolution (SR) has to be used to increase the quality of the eye images and to keep or improve the recognition performance. We propose an Efficient Single Image Super-Resolution algorithm, which takes into account a trade-off between the efficiency and the size of its filters. To that end, the method implements a loss function based on the Sharpness metric used to evaluate iris images quality.

Introduction

Smartphones, and mobile devices in general, play a central role in our society now-a-days. We use them on a daily basis not only for communication purposes, but also to access social media and for sensitive tasks such as online banking. In order to increase the security level of those more sensitive applications, verifying the subject's identity plays a key role. To tackle this requirement, many companies are currently working towards creating applications to verify the subject's identity by comparing a selfie image with the reference face image stored in the embedded chip of an ID-Card/Passport and a selfie image using Near Field Communication (NFC) from smartphones. This represents a user-friendly identity verification process, which can be easily embedded into numerous applications. However, this verification process also faces some challenges: for instance, that selfie image is captured in an uncontrolled scenario, where occlusions due to wearing a scarf in winter or a hygienic facial mask during a pandemic such as COVID-19 may hinder the performance of general face recognition algorithms.

Above mentioned reasons have increased the interest on periocular based biometrics. In particular, it has been shown that periocular images captured with mobile devices for recognition purposes are mainly acquired as selfie face images. In this work, we have a twofold goal: verify a biometric claim in a verification transaction from a smartphone selfie periocular image in the visible spectrum (VIS) and propose an efficient super-resolution approach.

Literature Survey

Face biometrics have attracted significant attention as a technology for secure access to mobile devices. This is because almost all smartphones have RGB cameras suitable for capturing faces, and the required user interaction is acceptable given the popularity of ‘selfies. Most of the traditional methods for face biometrics may not be amenable to native execution on mobile hardware due to their limited memory and computing power. Consequently, a number of algorithms specifically designed or adapted to the mobile environment have been proposed for face biometrics. However, the state-of-the- art related to face biometrics in a mobile environment is not well known. This paper thoroughly and critically surveys face biometrics in terms of face detection and normalization, recognition, and anti-spoofing methods proposed for mobile devices. The overall aim is to improve understanding and discuss the advantages and limitations of the existing methods. Further, challenges and future research directions are identified for further research and development.

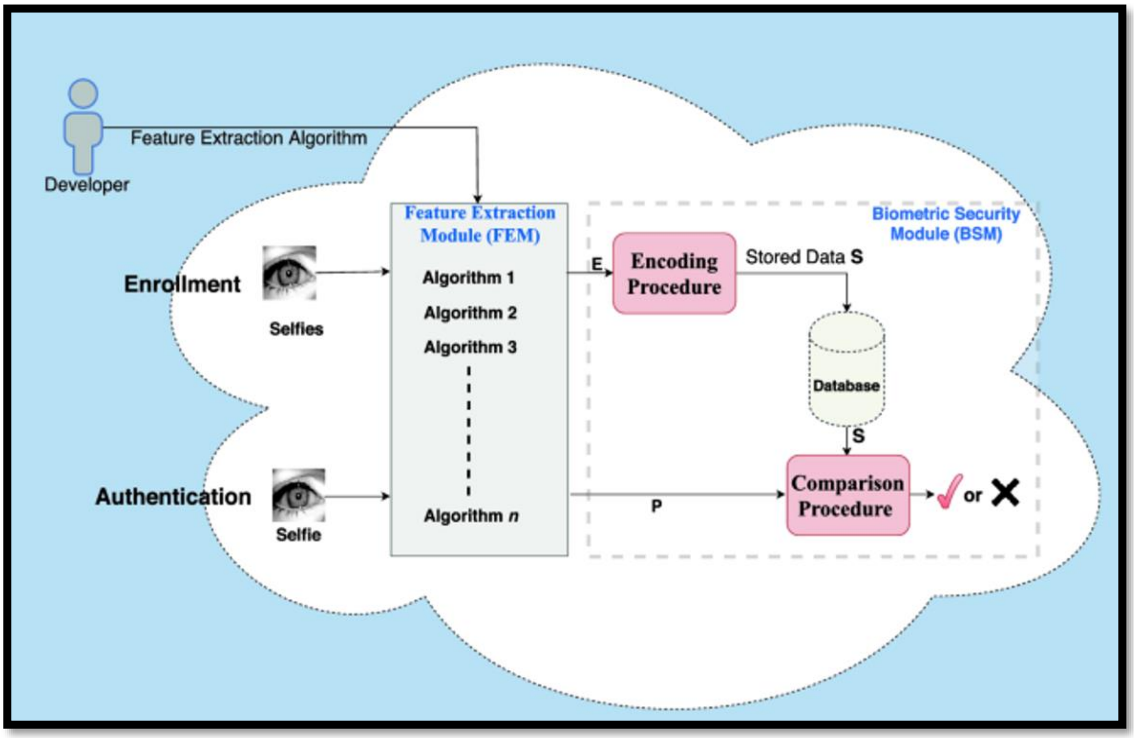
The general face recognition pipeline consists of the following steps:

(1) face image acquisition,

(2) face detection (determining if there exist a face in an image and if so segmenting it).

(3) face normalization and feature extraction.

(4) identity verification by matching features from two face images: a prior enrolment vs. the presented test query.



The performance of a typical face recognition system is affected by intra-class variations attributed to factors such as pose, occlusion due to prescription glasses, facial expression, make-up and illumination variations. However, existing methods for face detection and recognition may not be readily adaptable to mobile environment because of the following factors:

- Due to device mobility and operation in an uncontrolled environment, face images acquired using mobile phone's front facing cameras are usually degraded due to factors such as specular reflection, motion-blur, illumination variation and background lighting. Therefore, more efficient and robust methods may be required for integration in the mobile devices.
- Although the computational power of mobile devices is growing rapidly, it still may not be sufficient for real time operation of highly accurate and computationally costly face biometrics. A good user experience requires the whole face recognition process to take under one second. Given that about half of that time is spent by the camera module to initialize, meter, and capture the

image, an ideal face recognition module should take less than half a second.

The goal of this survey is to advance the state-of-the-art in mobile face detection, recognition and spoof detection by improving the understanding of the existing methods. To this aim, specific contributions of this paper are as follows:

- analysis of the existing methods on mobile face pre-processing, detection, and recognition through key attributes
- tabulation of the reported results from existing literature in terms of normalization time and reported accuracy
- discussion of the countermeasures against spoof attacks for face biometrics in mobile devices.

Mobile face detection and normalization methods:

Face detection is the first important step in face recognition. This is often implemented as a binary classification task in which a classifier trained on example faces decide whether a particular region of the image contains a face or not. In the context of mobile face detection, algorithms can be broadly classified into (a) skin-tone based, (b) machine learning-based, and (c) combination of skin-tone and machine learning-based.

Mobile face recognition methods:

After face detection and normalization, the identity of the subject is verified by classifying a test image as belonging to the same (genuine) or different identity (impostor).
The existing literature on mobile face recognition can be broadly categorized into (a) client-server based or (b) device based. In the client-server approach, face acquisition, face detection and sometimes feature extraction are performed on the device side. The remaining computationally intensive tasks such as classifier

Countermeasures for mobile face spoof attacks:

Vulnerability of biometric system to spoof attacks is well-known. Despite progress in anti-spoofing methods, face spoofing still poses a major threat to face recognition system. Compared to attacks against fingerprint, iris or speech recognition systems, the ubiquitous nature of image acquisition devices such as cameras and social medias, such as Facebook, allow attackers to acquire facial images of a user easily and discretely. A study conducted using commercial.

Challenges and future research directions:

One of the main challenges involve developing accurate and computationally efficient methods for face biometrics in mobile environment. Table 1, suggest maximum face detection accuracy of 99.7% and fastest processing time of 0.006 s. The average reported face detection accuracy is 89.0% and a processing time of about 5.56 s. Table 2 suggest maximum reported face recognition accuracy of 96.0% with processing time of 0.1 s. The average reported accuracy is 83.6% with 1.64 s of processing time
This survey suggests that existing mobile face detection methods can be categorized into skin-tone, machine-learning and those based on the combination of both. Similarly, mobile face recognition methods can be broadly categorized into client-server or device oriented. The average reported mobile face detection and recognition accuracies of 89% and 83%, respectively, are quite low for mobile-based user authentication. Current countermeasures to face spoof attacks use motion, texture, image.

This survey suggests that existing mobile face detection methods can be categorized into skin-tone, machine-learning and those based on the combination of both. Similarly, mobile face recognition methods can be broadly categorized into client-server or device oriented. The average reported mobile face detection and recognition accuracies of 89% and 83%, respectively, are quite low for mobile-based user authentication. Current countermeasures to face spoof attacks use motion, texture, image.

Literature review

Super-resolution (SR) is the process of recovering a high-resolution (HR) image from a low-resolution (LR) one. Supervised machine learning approaches learn mapping functions from LR images to HR images from a large number of examples. The mapping function learned by these models is the inverse of a downgrade function that transforms HR images into LR images. Such downgrade functions can be known or unknown.

Many state-of-the-art super-resolution models learn most of the mapping function in LR space, followed by one or more up sampling layers at the end of the network. Earlier approaches first up sampled the LR image with a pre-defined up sampling operation and then learned the mapping in HR space (pre-up sampling SR). A disadvantage of this approach is that more parameters per layer are required because they used more convolutional layers than small filters, which leads to higher computational costs and limits the construction of deeper neural networks. Sr requires that most of the information contained in an LR image must be preserved in the SR image. Sr models therefore mainly learn the residuals between LR and HR images. Residual network designs are therefore of high importance: identity information is conveyed via skip connections whereas reconstruction of high frequency content is done on the main path of the network

Dong et al proposed several SISR algorithms which can be categorized into four types: prediction models, edge-based methods, image statistical methods, and patch-based (or example-based) methods. This method uses 2 to 4 convolutional layers to prove that the learned model performs well on SISR tasks. We can conclude that using a larger filter size better than using deeper convolutional neural networks (CNN’s).

Kim et al. proposed an image SR method using a deeply recursive convolutional network (DRCN), which contains deep CNN’s with up to 20 layers. Consequently, the model has a huge number of parameters.

However, the CNN’s share each other’s weights to reduce the number of parameters to be trained, thereby being able to succeed in training the deep CNN network and achieving a significant performance. We ca conclude that deeper networks are better than large filters.

Yamanaka et al proposed a deep CNN with a residual net, skip connection and network (DCSCN) model achieving a state- of-the-art reconstruction performance while reducing by at least 10 times the computational cost. According to the existing literature, deep CNN’s with residual blocks and skip connections are suitable to capture fine details in the reconstruction process. In the same context, it is propose the pixel-shuffle and transposed convolution algorithm to extract the most relevant features from the images. The transposed convolutional layer can learn up-sampling kernels. However, the process is similar to the usual convolutional layer and the reconstruction ability is limited. To obtain a better reconstruction performance, the transposed convolutional layers need to be stacked, which means the whole process needs high computational resources. Conversely, pixel-shuffle extracts feature from the low-resolution images. But we can argue that batch normalisation loses scale information of images and reduces the range flexibility of activations. Removal of batch normalisation layers not only increases SR performance but also reduces GPU memory 40%. This way, significantly larger models can be trained.

Yu et al. Proposed the key idea of wide activation to explore efficient ways to expand features before ReLU, since simply adding more parameters is inefficient for smartphone-based image SR scenarios. The authors present two new networks named wide activation for efficient and accurate image super-resolution (WSDR). These networks (WDSR-a and WDSR-b) yielded better results on the large-scale div2k image super resolution benchmark in terms of PSNR with the same or lower computational complexity. Similar results but with a larger number of parameters are presented by LIM et al. In a model called enhanced deep residual networks for single image super resolution (EDSR).

Specifically for biometric applications, some papers have explored the use of SR in iris recognition in the visible and near-infrared spectrum. Ribeiro et al. proposed a SISR method using CNN's for iris recognition. In particular, the authors test different state- of-the-art CNN architectures and use different training databases in both the near-infrared and visible spectra. Their results are validated on a database of 1,872 near-infrared iris images and on a smartphone image database. The experiments show that using deeper architectures trained with texture databases that provide a balance between edge preservation and the smoothness of the method can lead to good results in the iris recognition process. Furthermore, the authors used PSNR and SSIM to measure the quality of the reconstruction. More recently, Alonso Fernandez et al presented a comprehensive survey of iris SR approaches. They also described an eigen-patches reconstruction method based on the principal component analysis and eigen-transformation of local image patches. The inherent structure of the iris is reproduced by building a patch-position-dependent dictionary. The authors also used PSNR and SSIM to measure the quality of the reconstruction in the NIR spectrum and in the NTNU database in the visible spectrum.

Periocular recognition based on traditional feature extraction methods such as intensity, shape, texture, fusion, and off-the-shelf CNN features with pre-trained models has been widely studies. However, to the best of our knowledge, only a few papers have explored the use of SR methods to improve the quality of the RGB images coming from periocular selfie captures.

Padole and Proença proposed a new initialization strategy for the definition of the periocular region-of-interest and the performance degradation factor for periocular biometric and the influence of Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP), Scale-Invariant Feature Transform (SIFT), Fusion at the Score Level, Effect of Reference Points of the eyes, Covariates, Occlusion Performance and Pigmentation Level Performance.

Raja *et al.* explore multi-modal biometrics as a means for secure authentication. The proposed system employs face, periocular, and iris images, all captured with embedded smartphone cameras. As the face image is captured closely, one can always obtain periocular and iris information with fine details. This work also explores various score level fusion schemes of complementary information from all three modalities. Also, the same authors used in [20] used in the periocular region for authentication under unconstrained acquisition in biometrics. They acquired a new database named Visible Spectrum Periocular Image (VISPI) and proposed two new feature extraction techniques to achieve robust and blur invariant biometric verification using periocular images captured by smartphones.

Ahuja *et al.* proposed a hybrid convolution-based model for verifying pairs of periocular RGB images. They composed a hybrid model as a combination of an unsupervised and a supervised CNN and augment the combination with SIFT model.

Hernandez-Diaz *et al.* proposed a method to apply existing architectures pre-trained on the ImageNet Large Scale Visual Recognition Challenge, to the task of periocular recognition. These networks have proven to be very successful for many other computer vision tasks apart from the detection and classification tasks for which they were designed. They demonstrate that these off-the-shelf CNN features can effectively recognise individuals based on periocular images.

More recently, Kumari and Seeja surveyed periocular biometrics and provided a deep insight of various aspects, including the periocular region utility as a stand-alone modality, its fusion with iris, its application in the smartphone authentication, and its role in soft biometric classification. In their review, the authors did not mention SR approaches.

Therefore, most of the proposed studies on mobile face biometrics have emphasized developing computationally efficient methods (low memory and CPU impact) for face detection and recognition. Face normalization reduce the effect of intra-class variations such as lighting and pose variation through pre-processing and registration routines. An efficient SR architecture need to be proposed, using only a feature extractor and one block based on recursive learning of reconstruction. And several others need has been reviewed and are to be done. Therefore, we have tried and worked on it, the main contributions from this work can be summarised as follows:

- A recursive pixel-shuffle technique is introduced over a transposed convolution to extract and keep fine details of periocular images.

- A novel loss function that includes the LoG sharpness iris quality metric and the SR loss function was proposed.

- A significant reduction of the number of parameters in comparison with the state-of-the-art is reported.

- A novel database for selfie periocular eye images was acquired and is available for researchers upon request.

- A periocular verification system based on an embedded vector from three pre-trained models with an SR-based pre-processing of the samples (x2, x3 and x4) was tested.

- A benchmark between deep learning approaches and a handcrafted method is reported.

- A full analysis of the influence of SR on selfie biometrics scenarios with traditional resizing methods was also included.

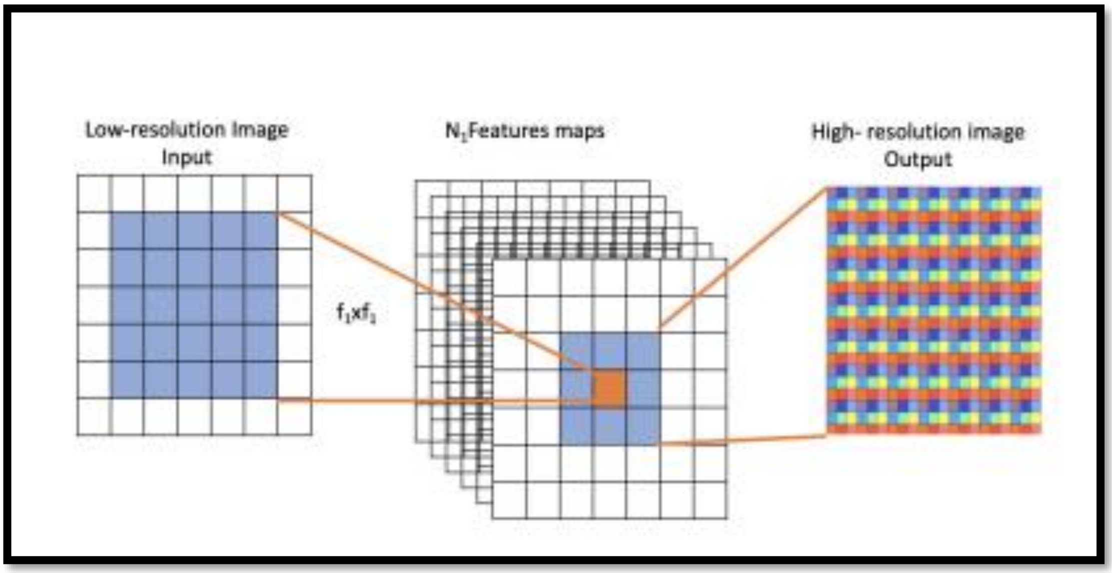
Implementation

1) Pre-processing

The original RGB images captured with a smartphone represent an additive colour-space where colours are obtained by a linear combination of Red, Green, and Blue values. The three channels are thus correlated by the amount of light on the surface. In order to avoid such correlations, all the images were converted from RGB to YCbCr. The YCrCb colour space is derived from RGB, and separates the luminance and chrominance components into different channels. In particular, it has the following three components: i) Y, Luminance or Luma component obtained from RGB after gamma correction; ii) $Cr = R - Y$, how far is the red component from Luma; and iii) $Cb = B - Y$, how far is the blue component from Luma. We only use Y component in this work because stored the high-resolution luminance information. Instead of CbCr that comprises the image information. The periocular image areas were automatically cropped from faces to the size of 250×200 pixels. The low-resolution version of the images is generated automatically in the training process using a resize function based on a bicubic interpolation to reduce the images to the half size for SR-X2, to the third part for SR-X3 and a quarter to SR-X4.

2) Feature extraction

As mentioned above, the Y component of the converted image is used as input for our model. Several patches of 32×32 and 48×48 pixels were extracted from the image and used to grasp the features efficiently. We look for the features that achieve a better trade-off between the number and size of filters of each CNN layer. Seven blocks of 5×5 and 3×3 have been selected after several experiments. The information is extracted using small convolutional blocks with residual connections and stride convolutions in order to preserve both the global and the fine details in periocular images. Only the final features from 3×3 and 5×5 pixels are concatenated, following the recursive pixel-shuffle approach (see Fig. 3). These local skip connections in residual blocks make the network easier to optimise, thereby supporting the construction of deeper networks. A model with transpose convolution instead of pixelshuffle was trained to explore the quality of the reconstruction images [7]. See Fig. 4. Transpose convolution operates conversely to normal convolution, predicting the input base



3) Reconstruction

Our reconstruction stage uses only one convolutional block with 2 layers (Conv + Relu + Conv) in a recursive path. This block includes 3 × 3 convolutions and pixel-shuffle algorithm (see Fig. 3) to create a high-resolution image from a low-resolution input. Batch normalisation was removed. An optimised sub-pixel convolution layer that learns a matrix of up-scaling filters to increase the final LR feature maps into the SR output was used.

4) Perceptual loss function

The ISO/IEC 29794-62 on iris image quality introduced a set of quality metrics, that can measure the utility of a sample. Based on the NIST IREX evaluation3 a sharpness metric was identified as strongly predictive for recognition performance. We follow this finding and measure: $LoG(x, y) = -\frac{1}{\pi\sigma^4}\left(1-\frac{x^2+y^2}{2\sigma^2}\right)e^{-\frac{x^2+y^2}{2\sigma^2}}$ (3) The Laplacian of Gaussian operator (LoG) is thus the sharpness metric used in this work. Calculation of the sharpness of an image is determined by the power resulting from filtering the image with a Laplacian of Gaussian kernel. The standard deviation of the Gaussian is 1,4

Now, it is important to highlight that the loss function aims to improve the quality of the reconstruction. To that end, we combine the SSIM and PSNR classical SR metrics with the sharpness metric for iris images recommended, as follows: L(ILR, IHR) = 0.5 · LoG (ILR, IHR) · [0.25 · SSIM (ILR, IHR) +0.25 · PSNR(ILR, IHR)] (4) where ILR represents a low-resolution image, IHR the corresponding high-resolution image recovered, and LoG the sharpness as defined in Eq. 3. The best values of the weights (w1, w2 and w3) for each specific metric (i.e., 0.25, 0.25 and 0.50) were estimated in a grid search with a train dataset.

5) Periocular recognition

Most traditional methods in the state-of-the-art are based on machine learning techniques with different feature extraction approaches such as HOG, LBP, and BSIF, or the fusion of some of them [8]. However, today we have powerful pre-trained deep learning methods based on facial images. Using transfer learning techniques, the information extracted from some layers using fine-tuning techniques or embedding approaches could be suitable to perform periocular verification. This is the approach followed in this work. This task involves information from periocular images estimating an eye embedding vector for a new given eye from a selfie image. An eye embedding is a vector that represents the features extracted from the eyes periocular images. This comparison occurs using Euclidean distance to verify if the distance is below a predefined threshold, often tuned for a specific dataset or application. For this paper, a VGGFace [17], FaceNet [16] and ArcFace [18] models have been used as a feature extractor for periocular recognition. Also, a comparison with BSIF handcrafted featured is included.

Conclusion

In this paper, we have proposed an efficient and accurate image super resolution method focused on the generation of enhanced eyes images for periocular verification purposes using selfie images. To that end, we developed a two-stage approach based on a CNN with pixel-shuffle, a new loss function based on a sharpness metric, derived from the ISO/IEC 29794-6 standard for iris quality, and a selfie periocular verification proposal. Overall, there are marginal improvements for verification systems when only the size of the images is considered in combination with SR images. The uncontrolled conditions such as sunlight, occlusions, rotations, or the number of people in an image when a remote selfie is captured could be more challenging than the image size for RGB selfie images.

In this research, SR helps in maintaining the recognition accuracy when selfies are captured at different distances. That is, in realistic scenarios in contrast to fully controlled conditions.

Future Work

In future work, we will continue to collect images to train a specific periocular verification system based on CNN from scratch and/or using transfer-domain techniques. Concerning the number of images, we believe that if we use state-of-the-art pre-trained models, the machine learning-based methods could be replaced by the CNN models. The selection of the pre-trained models should be taken into account.

References

- C. Dong, C. C. Loy, K. He and X. Tang, "Image super-resolution using deep convolutional networks", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295-307, Feb. 2016.

- J. Kim, J. K. Lee and K. M. Lee, "Deeply-recursive convolutional network for image super- resolution", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016.

- J. Yamanaka, S. Kuwashima and T. Kurita, "Fast and accurate image super resolution by deep CNN with skip connection and network in network" in Neural Information Processing, Cham, Switzerland:Springer, pp. 217-225, 2017.

- C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, et al., "Photo-realistic single image super-resolution using a generative adversarial network", *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 105-114, 2017.

- J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, et al., "Wide activation for efficient and accurate image super-resolution", *CoRR*, pp. 1-10, Aug. 2018.

- K. Ahuja, R. Islam, F. Barbhuiya and K. Dey, "Convolutional neural networks for ocular smartphone-based biometrics", *Pattern Recognit. Lett.*, vol. 91, pp. 17-26, May 2017

- K. Hernandez-Diaz, F. Alonso-Fernandez and J. Bigun, "Periocular recognition using CNN features off-the-shelf", *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, pp. 1-5, 2018.