

# Understanding low resolution facial image from image formation model

Haichuan Wang, h.wang@student.utwente.nl, s1705911

**Abstract**—A camera contains a planar, rectangle photosensitive sensor divided into square pixels and the RGB color is interpolated in the pixel when light reflect off the object traverse the camera lens and hit the sensor. All today’s digital images have gone through some kind of degradations and restorations during acquisition, coding, transmission and processing steps. This degradation and restoration disturb the original information of image from different level. Normally, high resolution image represents the original image better than the low resolution ones since it contains more discriminative information. Down sampling high resolution image gives the image are far more better quality than the real low resolution ones since many processes are playing a role in image formation procedure. For instance, the viewing angle is much smaller when the image recorded at a large distance, resulting geometric difference between the image captured closeby. In this paper, a camera model is developed to study and analyze the image formation procedure. Based on the model, all the parameters are carefully estimated by computing the minimal average difference in either spatial domain or frequency domain. The experimental results showed that the proposed method provide a valid but simple solution to narrow down the difference between different resolution facial images visually and numerically.

**Index Terms**—Image formation, PSF, Image sharpness, sensor noise, camera model

## I. INTRODUCTION

**B**IOMETRICS facial recognition is based on comparing the selected features with the given database and it has been a huge success for high resolution frontal facial image recognition in the past few years especially with deep learning. For instance, N Zhang, etc [5] proposed a method for facial recognition by using multiple cues with deep learning. However, deep learning requires a lot of low resolution image data that cannot be easily acquired. One simple solution is to obtain low resolution image data from down sampling the high resolution ones, but the realistic low resolution images differs a lot from the down-sampling ones. The most left image in Figure 1 is real low resolution image while the other four down-sampling are captured at different distance. As can be clearly found that the down-sampling images are still much better than the real low resolution ones.

There are many reasons related to digital camera model that make a difference between low resolution facial image and high resolution ones. Most of the existing works assumes the real low resolution image can be obtained by adding additive white Gaussian noise to the down-sampling blurred RGB image. For example, Sumit Shekhar, etc [4] proposed a Synthesis method for low res face recognition with such kind of image degradation and restoration model. However, the type of noise generated by today’s digital camera is much

more complex, depending on camera brand and series, as well as camera settings (ISO, shutter speed, aperture, sharpen and etc). In addition, the imaging system may also introduce some distortion and artifacts in signal due to defocused lens and some other reasons.

A typical camera model is the ideal pin hole model, which describes the mathematical relationship between the coordinates of a point in three-dimensional space and its projection onto the two dimensional image plane, where the camera aperture is described as a point and no lenses are used to focus light. However, the pin hole model does not take geometric distortions and blurring of unfocused objects into consideration, which might be a good estimation for high quality image but not low res ones. There are quite some geometric distortion, unfocused objects and noise in low res images. What kind of camera model can better explain how the 3d object is mapping onto the 2d image plane for low res images? How can we apply the camera model to perfectly reconstruct the low res facial image from high res ones? A new camera model is introduced in this paper to study and analyze the image formation procedure.

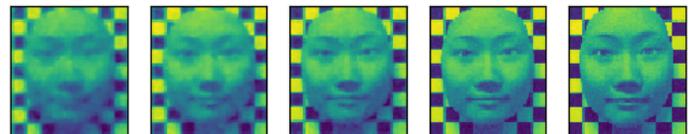


Fig. 1: Down-sampling image at different recording distance

### A. paper organization

The rest of paper is arranged as follow. The Section II describes the image formation and degradation model. The experiments and measurements of model are discussed in Section III. Finally, the IV concludes the paper with short summary and discussion.

## II. IMAGE FORMATION WITH CAMERA MODEL

An image is formed on image plane of the camera and then measured electronics in terms of RGB pixel numbers. One typical model describing this procedure is the ideal pin hole model, where only light source from scene that pass the camera aperture can drop on the image plane. However, this

model is far from the truth in reality and a improved version given in Figure 2 [1] shows how the image formation works in today's digital camera.

This paper is mainly focus on unavoidable factors appearing in camera model and all the setttable factors such as camera brand and setting, all the test images are captured with Casio EX-FC100 camera under same aperture,shutter speed and ISO setting. In addition,the white balance is set to automatic mode to compensate the light illumination. Moreover, a mask placed on chessboard is captured with different capture distance camera. The mask is used to guarantee that the captured facial image are frontal and expressionless.And the chessboard behind of the face is very useful for selection of region of interests and image alignment.

However, the noise caused by optics and sensor can not be removed with camera setting and low res image tends to become more noisy. The type of noise also differs quite some between the mostly used sensor CMOS and CCD. Besides, the low resolution facial tends to have more optical blur and lens defocus, which can be modeled as the convolution of original image with some blur kernel. And the image sharpness and quantization error(A/D) becomes less obvious for high res image after down-sampling.

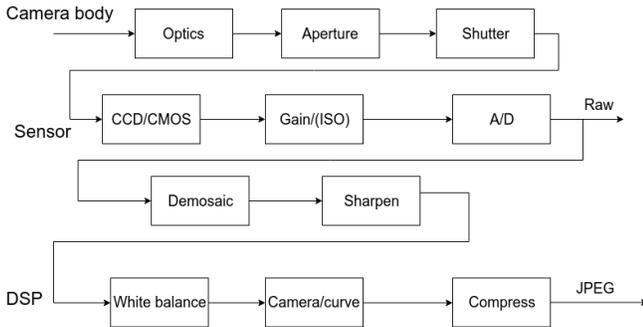


Fig. 2: The image formation camera model

### A. Point spread function

The image taken by today's digital camera is always blurred in a certain degree and the blur is convolved with the incoming image to produce the point-sampled image. The shape of convolution kernel, which is known as PSF(Point spread function) or OTF(optical transfer function) is caused by several factors, including lens blur and radial distortion, anti-aliasing filter in front of the CMOS/CCD sensor, as well as the finite image plane area[1]. The equation 1 describes the incoming image  $f_{n,m}$  convoluted with PSF kernel  $h_{k,l}$  and the motion blur kernel  $h_{k,l}$  is assumed to be Gaussian distribution.

$$g_{m,n} = \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} h_{k,l} f_{n-k,m-l} \quad (1)$$

Fourier transform is a very helpful tool to analyze the frequency characteristics of a filter kernel or an image. The equation 2 describes how to convert spatial image into frequency

domain and  $f(m,n)$ ,  $F(k,l)$  are spatial image and frequency domain image respectively.

$$F(k,l) = \sum_{m=0}^M \sum_{n=0}^N f(m,n) e^{-i2\pi(\frac{km}{M} + \frac{ln}{N})} \quad (2)$$

The left two images in Figure 3 are low quality spatial image and its corresponded frequency domain one while the right two are that of high resolution ones. As shown in the Figure 3, the high quality one at the right is much better quality than the left one. Besides, the relative size of the nose and eyes seems to be larger than the left one. As seen from the frequency domain image, the image contains all frequency components, but the magnitude of high frequencies is much smaller. Therefore, low frequencies contain more image information than the higher ones. And there are two dominating directions stripes in the Fourier domain image, one passing horizontally and one vertically through the center of origin [8]. And the central stripes of the Fourier transform spectrum has brighter and thicker pixel intensities. The right image tends to have more high frequencies components spread along with the two central stripes, which is mainly due to the chessboard at the back. The blur kernel(PSF) of Gaussian distribution can be applied to the high res facial image to get more blurred low res ones.

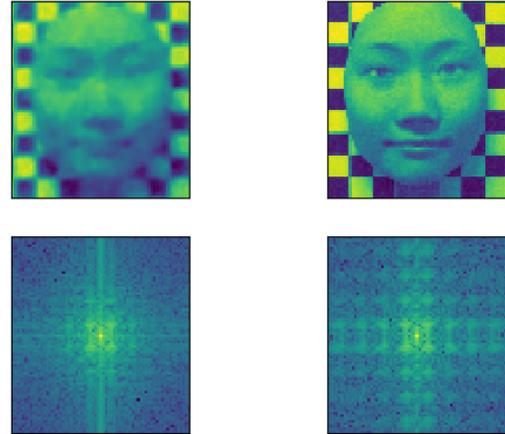


Fig. 3: Different resolution image in spatial domain and frequency domain

In order to optimize OTF(optical transfer function), the Gaussian kernel  $h$  with tunable standard deviation is applied to down-sampling image until the minimal difference between real low resolution image is found. The equation 3 shows the mathematical expression of the state-of-art, where  $o$  is real low resolution image,  $f$  is high resolution image,  $D$  is the down-sampling operator[6].

$$D_{average} = \arg \min_h ||o - D(h * f)|| \quad (3)$$

### B. Image noise

Noise always appears in digital images during image acquisition,coding,transmission and processing steps. Noise tells

unwanted information and always produces undesirable effects in digital image such as artifacts,unrealistic edges,unseen lines,corners,blurred objects and disturb background scene[7]. Digital noise may occur differently from various kind of source such as Charged Coupled Device(CCD) and Complementary Metal Oxide Semiconductor sensors. The table II-B lists all noise source from CCD and CMOS sensors. Besides, additive white Gaussian noise may occur in the digital image due to the mimic effect of many process in nature such as optics and electronics. The white Gaussian noise is available for all frequencies, but the low res image tends to have more white Gaussian noise.

sensor	noise	model	importance
CMOS	shot noise	Poisson	++
	quantization noise	uniform	++
	thermal noise	Gaussian	++
	dark noise	-	+
CCD	shot noise	Poisson	++
	quantization noise	uniform	++
	readout noise	Gaussian	++
	dark noise	-	+
optics	White noise	Gaussian	+

Even though there exists some different kinds of noise in CCD and CMOS imaging sensors, the noise model is more or less similar. In the paper, the image is captured with CMOS imaging sensor. As for CMOS sensor, there are mainly shot noise,thermal noise,quantization noise and dark noise. The thermal noise caused by poor illumination and high temperature of sensor and connected electronics is additive and independent Gaussian distribution [7]. The Gaussian noise is as

$$N_{thermal}(g) = \frac{1}{\sqrt{2\pi} * \sigma^2} e^{-\frac{(g-\mu)^2}{2\sigma^2}} \quad (4)$$

where  $\sigma$  is the standard derivation and  $\mu$  mean value. The shot noise is mainly caused by random fluctuations of photons.The photon shot noise follows Poisson distribution and is given below

$$N_{shot} = \frac{\lambda^k e^{-\lambda}}{k!} \quad (5)$$

The quantization noise is due to the quantization error of ADC(analog-digital-converter), which is approximately uniform distributed. Due to the down-sampling, the quantization noise of high quality image becomes smaller. The distribution is given as

$$N_{quantization}(g) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leq g \leq b \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The dark noise, depends on the flow of current over exposure time. When the exposure time of two images is same, the noise difference is very trivial. By manually set the camera exposure time to same, the dark noise difference between images can be neglected.

### C. Image sharpness

The image sharpness describes derivative of brightness with respect of space. Due to the down-sampling, the image sharpness of high resolution facial image becomes less evident,which can be clearly seen in frequency domain. The Difference of Gaussian(DoG) is a feature enhancement algorithm that can be used to increase the visibility of edges and other detail present in the digital image. The DoG,as band pass filter,removes low and high frequency present in grayscale image. The equation of DoG is given as

$$F = I * \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma_1^2}} - I * \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma_2^2}} \quad (7)$$

In the equation, $I$  is the original grayscale image,  $F$  is the DoG image. The image sharpness is modeled as the summation of the original image with DoG image, which is given as

$$F = I + \alpha * DoG(I) \quad (8)$$

In the equation, $I$  is the grayscale image,  $F$  is unsharpened image.

### D. Geometric distortion

The angle of view of camera is adjusted automatically and the relative size of object appears to be different when the camera is towards or away from the camera. One of the most direct noticeable feature is that the relative size of object seems to be larger in image recorded nearby, which can be also clearly seen from Figure 1. As can be seen down-sampling facial image, the nose seems to be larger than that of real low res ones. More importantly, there is also some geometric difference due to different recording distance.

In summary, there are many processes involved in the image formation model, but there are four major effects between low res facial image and high res ones, namely PSF, image sharpness,noise and dolly effect. The quantization noise and shot noise does not make much difference between different res image, which thus can be simply ignored. The thermal Poisson noise can be as a limit of Gaussian distribution when  $k$  is very large. Therefore, the total noise can be seen as Gaussian distributed noise. The dolly effect is very hard to model and analyze and the impact is assumed to be relatively small. The Figure 4 shows the overall approach to generate the low resolution facial image from high res ones. The first step is, of course, down sampling the high res image to same size. The second step is to apply Gaussian PSF blur to the down-sampling image. After that, the output image is added with its Parameterized DoG. Finally, compare the FFT difference in frequency domain with Gaussian noise. There are six parameters to be adjusted, namely standard deviation  $s$  of Gaussian blur kernel of PSF,standard deviation  $s_1, s_2$  amplitude of DoG for image sharpness and standard deviation  $s$  and amplitude  $a$  for Gaussian noise.

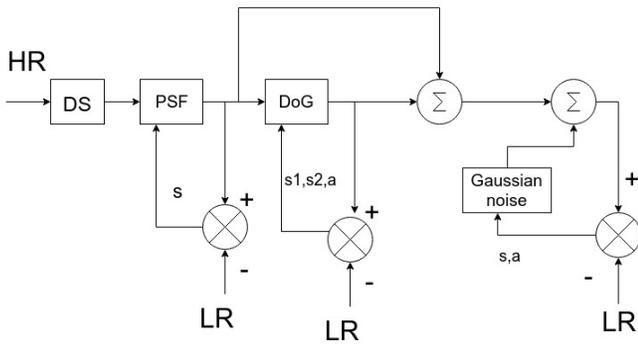


Fig. 4: The overall image formation method

The low res facial image, as a reference feedback, is used to tune the parameters. By minimizing the average difference between reconstructed image and real low res ones, the optimal parameters can thus be obtained.

### III. EXPERIMENT

In this research, all the image formation experiments are carried out with OpenCV (Open Source Computer Vision) Libraries on Ubuntu 16.04 platform. OpenCV is released under BSD (Berkeley Software Distribution) license and hence it is free for both academic and commercial use. It has C++, java and python interfaces that supports Linux, Windows, Mac OS, iOS and Android. All the experiment and measurement results are generated from python libraries of OpenCV and others. The mask fixed on the chessboard is captured with moving casio EC-FC100 digital camera. The recording distance is unknown for all the facial image, so the pin hole mathematical model is used to estimate the distance ratio with the lowest resolution ones.

#### A. Point spread function

The PSF can be regarded as original image convolved with Gaussian kernel, so amount of PSF blur only depends on the standard deviation of Gaussian kernel. The optimal blurring parameter is acquired by minimizing the absolute average difference with the real low res image. The figure 5 shows the absolute average difference with different standard deviation. When the standard deviation goes up, the absolute average difference first drops and then increases. The minimal average image difference standard deviation parameter, regarded as the optimal blur of low quality image, is found at the bottom of the plot.

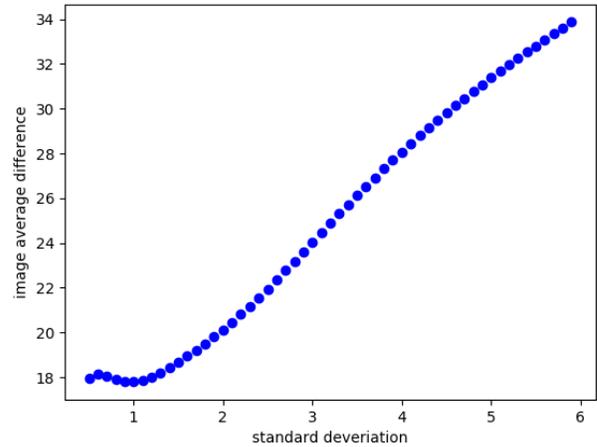


Fig. 5: The image difference with different standard derivation of PSF

The Figure 6 shows the low res image and reconstructed image in both spatial domain and frequency domain. Compared with the original down-sampling, the right spatial image becomes much blurred. As for Fourier domain image, both the vertical and horizontal stripes becomes more dispersed. But in general, the output image look more similar to the left real low res one.

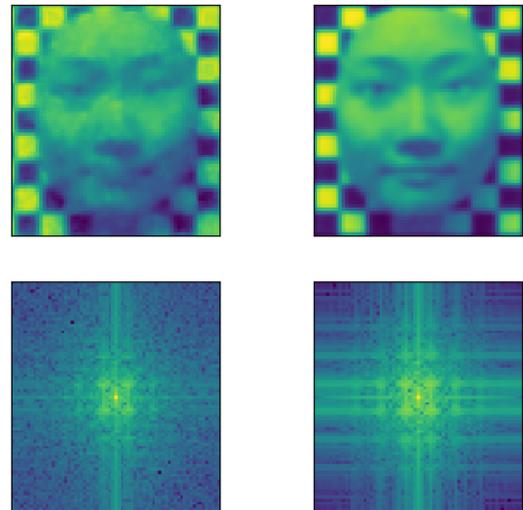


Fig. 6: image difference with different distance using PSF

The Figure 7 shows the optimal with different estimated distance ratio. The distance ratio, as discussed previously, is approximated with pin hole model. It is clear that the absolute average difference drops with some small fluctuation when the distance ratio goes up. It is much easier minimize the average difference with real low image from relatively low quality ones rather than high quality ones.

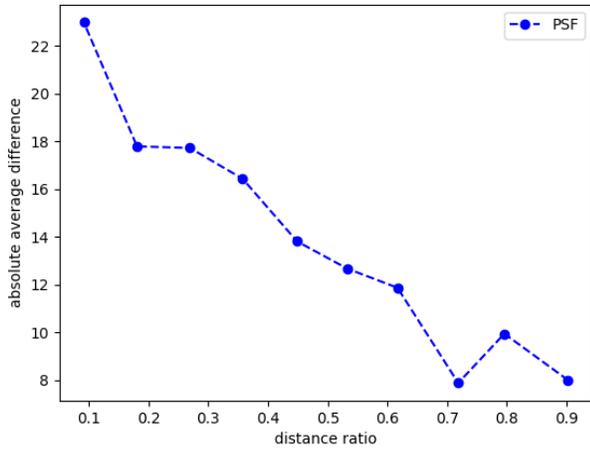


Fig. 7: The optimal average image difference with different distance ratio using PSF

B. image sharpness

The image sharpness describes the derivative of image pixel value with the respect of space. Due to down sampling, the image sharpness of high res ones becomes less evident. The Difference of Gaussian is the difference of output of two Gaussian filter with different amount of blur. The standard deviation  $\sigma$  represents the amount of blur and a bigger *sigma* will always gives bigger amount of blurring. The DoG can be easily acquired from subtracting one Gaussian blurred image from another.

The sharpness of low res image can be obtained from the summation of DoG image with the original ones. The two different standard deviation and the amplitude of DoG are unknown, which is to be determined with minimum average image difference in spatial domain. Estimating three parameters at one computation is computationally costly and it is also error prone due to floating number of CPU. The two parameters are computed with one random estimated parameters until the minimal average difference is found. Afterwards, repeat the process again with one of previously approximated parameters. The Figure 6 shows the average image difference with respect of the two different standard deviation when the amplitude of DoG is 0.8 and it can be clearly seen that the absolute average difference fluctuates with the standard deviation. The optimal standard deviation can be estimated at the valley of the image average difference and there might exist more than one minimal absolute average difference, which is mainly due to the floating number of CPU. The optimal standard deviation is estimated with any of the results first and then use the optimal standard deviation to determine the amplitude of DoG. Afterwards, a more precise amplitude is used to determine the standard deviation once more.

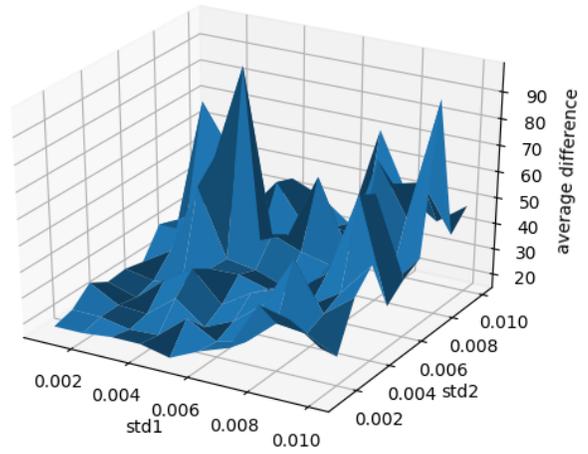


Fig. 8: The average image difference with different standard derivation using PSF and DoG

The Figure 9 shows the low res image and reconstructed ones with PSF and DoG in both spatial and Fourier domain. The amount of bright horizontal and vertical stripes in right FFT image is reduced compared with Figure 6.

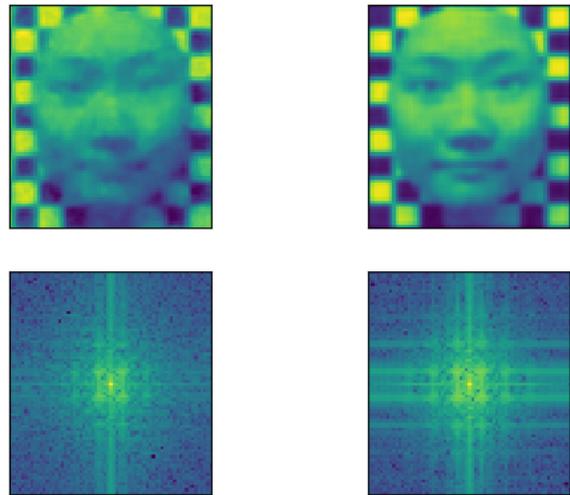


Fig. 9: Different quality image in spatial domain and FFT domain using PSF and DoG

The Figure 10 illustrates the absolute average image difference for image with and without DoG. Compared image only with PSF blur, the average difference drops slightly. Besides, the distance ratio still plays a very important role for the absolute average image difference.

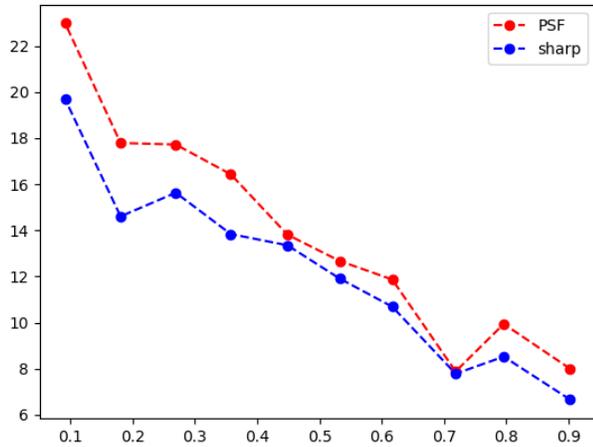


Fig. 10: The optimal average image difference with different distance ratio using PSF and image sharpness

In short, the image sharpness techniques does help to reduce the image absolute average difference, so the optimal DoG parameters selection is quite crucial.

C. Noise

Noise is random variation of brightness or pixel values in images and noise always produce unwanted effect in images. The noise can be analyzed in both spatial and Fourier domain. In spatial domain, the image is processed as it is and the pixel value changes with respect of scene. However, the Gaussian noise is independent of each pixel and subtracting one noise with another may enlarge the noisy pixel. In frequency domain, the image is analyzed with the change of pixel value in spatial domain and the white Gaussian noise has an inherently flat frequency spectrum. The Gaussian noise should be analyzed in frequency domain rather than in spatial domain. And the DFT(discrete Fourier transform) is a specific form of converting the spatial image into Fourier domain image and FFT(fast Fourier transform) is an efficient way of implementing DFT. Adding Gaussian noise will add noise in all frequencies, which will introduce some high frequency noise to the reconstructed image, so a Gaussian low pass filter is used to remove the high frequency components of Gaussian noise. The inverse fast Fourier transform(iFFT) can convert the Fourier image back to spatial image.

In Fourier domain, the difference between reconstructed DoG image and real low resolution image is computed, which is to be compared with low pass filtered Gaussian noise. The figure 11 shows the average FFT difference with the respect of different Gaussian noise standard deviation and low pass filter standard deviation. And the optimal standard deviations can be acquired at the valley of Figure. The spatial domain noise can be rebuilt with the estimated standard deviations and add the noise to the constructed image previously will get the final image, which is shown in Figure 12.

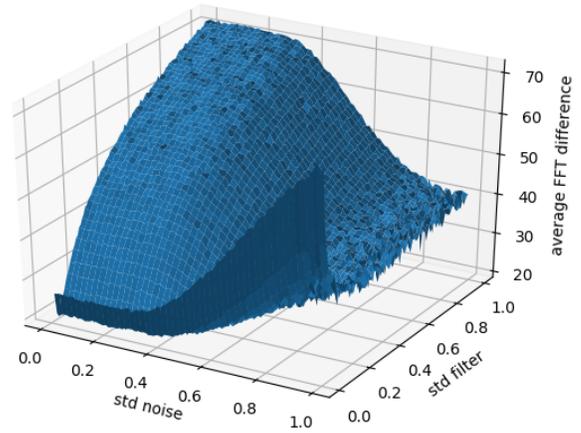


Fig. 11: The average FFT image difference with respect of std of Gaussian noise and low pass filter

The reconstructed facial image shown at the right of Figure 12 is very close to the real low resolution. However, the eyes and nose of reconstructed facial image still looks slightly bigger than that of the left one, which is due to the capturing distance during image recording.

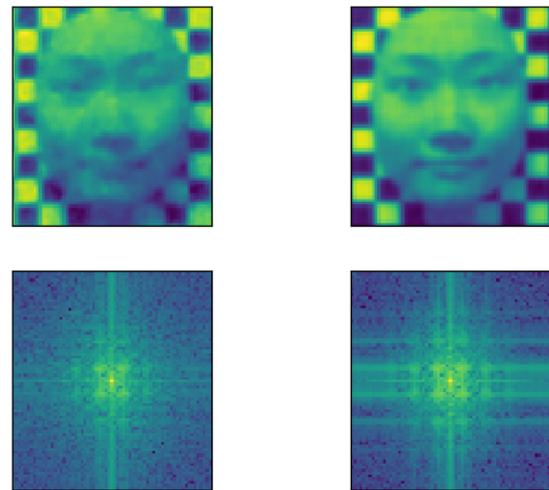


Fig. 12: The optimal image with Gaussian using PSF,DoG and noise

By applying same techniques to different distance recorded image, the Figure 15 is acquired. The facial image still differs quite some especially between lowest resolution ones and reconstructed high resolution ones, which is caused by many reasons. The most obvious cause is due to the dolly effect of camera when recording at different distance. The second reason is that the noise is much more complex than just Gaussian.

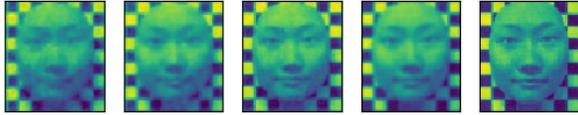


Fig. 13: final image

IV. ANALYSIS AND DISCUSSION

The paper shows a valid way to reduce the difference of image using the camera model. After applying the PSF, DoG and Gaussian noise to the down-sampling image, the average difference drops quite significantly. More importantly, the reconstructed image looks very similar to the real one especially when the image recording distance ratio is very large. However, the relative landmark size on the face differs quite a lot, which can be clearly seen in the Figure, so further research with geometric distortion is needed in order to perfectly synthesis the real low res images from the high res ones.

APPENDIX A

THE ORIGINAL DOWN-SAMPLING IMAGE DATASETS

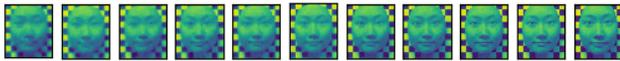


Fig. 14: original image dataset

APPENDIX B

THE IMAGE DATASET WITH IMAGE FORMATION TECHNIQUES

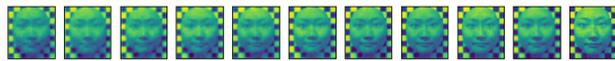


Fig. 15: image processed with camera model

ACKNOWLEDGMENT

The work is partially supported by Data Science of University of Twente, the Netherlands. The author would love to thanks for the group offer image data and equipments for the experiments and measurements.

REFERENCES

[1] Richard Szeliski, Computer Vision Algorithm and Applications 2010  
 [2] Milan Sonka, Vaclav Hlavac, Roger Boyle, Image Processing, Analysis, and Machine Vision, Fourth Edition  
 [3] Min-Chun Yang, Chia-Po Wei, Yi-Ren Yeh, Yu-Chiang Frank Wang, Recognition at a Long Distance: Very Low Resolution Face Recognition and Hallucination, ICB 2015  
 [4] Sumit Shekhar, Vishal M. Patel, Rama Chellappa Synthesis-based Robust Low Resolution Face Recognition  
 [5] Ning Zhang, Manohar Paluri, Yaniv Taigman, Rob Fergus, Lubomir Bourdev Beyond Frontal Faces: Improving Person Recognition Using Multiple Cues

[6] Joshi, Szeliski, and Kriegman PSF Estimation using Sharp Edge Prediction, 2008  
 [7] A REVIEW PAPER: NOISE MODELS IN DIGITAL IMAGE PROCESSING, Ajay Kumar Boyat and Brijendra Kumar Joshi  
 [8] Robert Fisher, Simon Perkins, Ashley Walker and Erik Wolfart, Hypermedia Image Processing Reference  
 [9] Yuxi Peng, Luuk Spreeuwers and Raymond Veldhuis, Designing a Low-Resolution Face Recognition System for Long-Range Surveillance, International Conference of the Biometrics Special Interest Group, 2016