Analysis of feature representation in dictionary learning and sparse coding algorithms for low resolution image

Suit Mun Ng 1 and Haniza Yazid 2

1, 2 School of Mechatronic Engineering, Universiti Malaysia Perlis (UniMAP) Kampus Pauh Putra, 02600, Arau, Perlis.
E-mail: ngsuimun95@gmail.com

Abstract. Super-Resolution (SR) is used to recover a high-resolution (HR) image from the image with low-resolution (LR). SR is important in the biometric identification and the face recognition is an area that bring attention to people. However, the performance of the current systems is affected by the resolution of the input images. Thus, this paper is focusing on the analysis of feature representations in dictionary learning and sparse coding methods for LR image. The input image is the Lena image in grey scale. A total number of 23 features were extracted from the image patches to develop different learned dictionaries using the k-singular value decomposition (k-SVD) algorithm. The denoised images were then produced by using the Douglas-Rachford algorithm. Most of the feature representations were able to produce a final image with Peak-to-Signal Noise Ratio (PSNR) and Structural Similarity Index Matric (SSIM) values of approximately 29 dB to 30 dB and 0.8300 to 0.8600 respectively. However, the denoised image produced with gradient direction obtained only 27.6676 dB and 0.7881 for PSNR and SSIM. Therefore, when different features were extracted for conducting the dictionary learning and sparse coding algorithm, denoised image with different PSNR and SSIM were produced at the end of the process.

1. Introduction
SR is defined as a method used to acquire more information about an image by recovering a HR image from one or multiple LR images without losing high frequency details [1]. Generally, LR image is generated from the corresponding HR image with the processes such as blurring, down-sampling or additive Gaussian noise. However, due to the physical boundaries of the related imaging devices, algorithm based solution has replaced the hardware solution in terms of improving the spatial resolution of the LR images [2]. Hence, a single-image SR reconstruction system based on dictionary learning and sparse coding by using different feature representations which extracted from the LR image patches is proposed in this paper. Essentially, SR is important in many applications in real life especially in the field of biometric information identification which includes the resolution improvement of faces, fingerprints and iris images [3]. Face recognition is known as one of the important areas that employed biometric identification. But unfortunately, the performance of the current face recognition systems is affected by the LR problems on the captured face images. These problems may occur as a result of hardware limitation, motion blur, illumination and movement when taking the images. Therefore, in this paper, the main focuses is the analysis of feature representations from the LR biometric image patches with the use of dictionary learning and sparse coding method in order to generate a satisfactory HR image.
2. Literature Review

Recently, SR is a procedure to find one or more HR images from one or more LR inputs information. The basic knowledge of SR is to produce a HR image by combining the non-redundant information contained in LR images [3]. The application of SR has also been applied in many real-world systems for example the satellite and aerial imaging [4], [5], medical image processing [6], facial image analysis, text image analysis [7], [8], sign and number plates reading, and biometrics recognition [5]. Then, SR can be differentiated into two types: single-frame and multi-frame based on the input LR materials. Mostly, single-image based SR algorithms work better than the multiple-based methods due to a single-image SR can be conducted by using limited information especially when multiple LR images are not available, in addition, multiple-image based methods are unstable as it is highly dependent on the approximation accuracy of the indications between the LR observations. Therefore, algorithm with single-frame is chosen to be used instead of the multi-frame. Then, Yue et al. in [3] stated that the practicability of the SR methods implemented by other researchers is still limited by the comparatively poor performance and time consumption problems. In [9], Tai and Ma described a comprehensive review related to the SR image and video reconstruction methods implemented by the researchers and also emphasised the future developments in the field of SR image reconstruction.

Next, the dictionary learning and sparse coding algorithms have also been discussed by numerous researches. The paper done by Zhang et al. summarized and presented the current sparse representation algorithms and discussed the inspiration, mathematical representations and further applications of these algorithms [10]. Furthermore, Wang et al. proposed a new model for SR image reconstruction by combining the advantages of sparse coding and deep network [11]. In [6], Jiang et al. proposed a single CT-image SR reconstruction scheme and they found that the performance of the denoised HR image improved after several operational iterations. In this paper, the k-SVD algorithm is used to create the overcomplete dictionary for sparse coding as referred to Aharon et al. in [12]. Then, Li and Qi solved the compressed sensing (CS) image recovery problem by using an algorithm called nonlocal Douglas-Rachford (NLDR) [13]. Hence, the Douglas-Rachford algorithm is also proposed to be used as the sparse coding method for denoising purposes.

3. Methodology

The analysis of feature representations extracted from LR images patches with the use of dictionary learning and sparse coding algorithm to obtain the HR image is discussed. Fig. 1. shows the system architecture for the proposed algorithm and the highlighted part indicates different features will be used in order to discuss the effectiveness of the feature representations in producing a HR image with good performance.
3.1. LR image construction

The HR image used in this paper is the images of Lena [14] in grey scale which obtained from the standard database. The HR image is saved in TIFF format with an image resolution of 512x512 pixels. Firstly, the HR image, $x$ was cropped and down-sampled into 256x256 pixels. Then, LR image, $y$ was produced by adding random Gaussian noise, $v$ as shown by Equation (1).

$$ y = x + v $$

3.2. Random Patch Extraction

Then, the LR image, $y$ produced is used to extract a large number $m$ of image patches, $Y$ with size of 10x10 pixels as described by Equation (2). The image patches, $m$ was calculated based on the number of atoms, $p$ in the dictionary by using the formula as shown in Equation (3).

$$ Y = \{y_j\}_{j=1}^{m} \in \mathbb{R}^{nxm} $$

$$ m = 20p, \text{where } p = 2n $$

After that, Equation (4) is used to estimate the random patch location with a total amount of $q$ Before extracting the images patches, $y_j$ from LR image, $y$.

$$ q = 3m $$
3.3. Feature Extraction

From the previous stage, the matrix \( Y \) was used to store the image patches, \( y_j \) extracted. In this case, the mean was first removed from the image patches, \( y_j \). Then, in order to analyse the features which are able to produce a HR image with good denoising result, a total number of 23 features was extracted from the image patches, \( Y \). These features includes: energy, gradient magnitude, gradient direction, first derivative, second derivative, standard deviation, maximum value, minimum value, variance, gradient mean, standard deviation of gradient, gradient variation, sum, smoothness, kurtosis, skewness, mean density, intensity variation, image contrast, local entropy filtering, local range filtering, local standard deviation filtering and entropy. The values obtained will be sorted in descending order to make sure that only the largest values were kept from the image patches, \( y_j \). Thus, an initial dictionary, \( D_0 \) with zero mean and unit norm was computed in this step.

3.4. Dictionary learning Algorithm (k-SVD)

Next, the k-SVD algorithm was applied in order to obtain the final learned dictionary, \( D \). Fundamentally, k-SVD is a technique which substitutes between sparse coding of the examples based on the initial dictionary and also a method of updating the dictionary atom to fit the data in a better condition [12]. The general function of k-SVD is shown in Equation (5) where \( Y \) is the matrix with all the known examples, \( D \) is defined as the final dictionary, \( X \) is the matrix of sparse representation vectors, \( x_i \) and \( k \) is the limit of sparsity (\( k \) is set to be 4 in this paper) [10].

\[
\arg \min_{D,X} \left\{ \| Y - DX \|_F^2 \right\} \text{ s.t. } \| x_i \|_0 \leq k, \ i = 1, 2, ..., N
\]

(5)

In this case, the sparse coding operation used to compute the sparse representation vectors, \( x_j \) during the process of k-SVD algorithm is called the Forward-Backward Iterative Scheme or Projected Gradient Descent. After that, the update of dictionary one column at a time is conducted by k-SVD algorithm. By doing this, the minimization error can be decreased per update of each dictionary atom and the sparse representation coefficients, \( x_j \) also changed with the update. Hence, a learned dictionary, \( D \) was produced at the end of the dictionary learning process.

3.5. Denoising by Douglas-Rachford Algorithm

Before proceeding with the image denoising step by using Douglas-Rachford algorithm, a large number \( m \) of random image patches \( Y \) was extracted again from the LR image, \( y \). This is to ensure that the random image patches used to find the sparse representation coefficient, \( x_j \) in denoising step is different with those patches used for developing the final dictionary, \( D \). Also, the mean was also removed from the image patches since the learned dictionary was having zero mean and unit norm.

Douglas-Rachford splitting method was formerly used to solve matrix problem as stated in [15]. This algorithm is used to obtain the sparse representation coefficients, \( x_j \) as the sparse coding problem can also be solved in an easier way by written it using a proximal splitting scheme. In this way, the auxiliary variable \( u = Dx \in \mathbb{R}^n \) is explained in Equation (6), provided that the \( F(x, u) \) and \( G(x, u) \) are also defined by Equation Error! Reference source not found. and Equation Error! Reference source not found. respectively.

\[
\min_{z} \{ (x, u) \in \mathbb{R}^p \times \mathbb{R}^n : F(z) + G(z) \}
\]

(6)

Then, the proximal mapping \( \text{Prox}_{\gamma F} \) and \( \text{Prox}_{\gamma G} \) as stated in Equation (7) and Equation (8) respectively was computed by using the convex functions \( F(z) \) and \( G(z) \). Lastly, Equation (9) shows the Douglas-Rachford iterations and the sparse representation coefficients, \( x_j \) were successfully computed after conducting the iterations.
3.6. Part Averaging
With the use of sparse representation coefficients, \( x_j \) computed from the previous step, the denoised patches, \( \tilde{y}_j = D x_j \) can be calculated. After calculating the denoised patches, \( \tilde{y}_j \), the mean which removed in the first place was inserted back to it. Lastly, the final HR image, \( f \) was also formed by using patch averaging formula as shown in Equation (10), where \( W_t \) is defined as the number of image patches that overlap in a given location \( t \).

\[
f(t) = \frac{1}{W_t} \sum_j \tilde{y}_j (t - a_j)
\]  

(10)

3.7. Evaluation
The final HR images produced by using different features extracted from the LR image were compared and analysed in order to evaluate the best feature which is able to create a biometric image with the best resolution. Therefore, the denoised images produced was evaluated by using the image quality evaluation indexes which are known as the PSNR and SSIM values. The reference image used to calculate the PSNR and SSIM values in this case is the original input HR image.

4. Result and Discussion
In this section, the results obtained throughout the process will be shown. First of all, the LR image produced by using the grey scale image of Lena obtained from standard database is shown in figure 2(C). This LR image produced was used to extract a 12000 images patches with size 10x10 pixels which stored in the matrix \( Y \). As explained before, the mean of each patches was removed to produce a dictionary with zero mean. Then, a total number of 23 features were extracted from these image patches. In this case, the values obtained for each feature were sorted in descending order and the values for the first 4000 image patches were kept to be used for the dictionary learning step later.

![Figure 2.](image-url)

Since the properties of each feature is different, therefore, the initial dictionary and the final dictionary developed by using the k-SVD algorithm was not the same. Fig. 3, shows one of the examples
result for the initial dictionary and learned dictionary obtained using the feature values which extracted based on the image contrast of the LR image patches.

![Initial and Final Dictionary](image)

**Figure 3.** (Left) Initial dictionary and (Right) final dictionary produced by feature image contrast

After that, a total number of 15876 image patches were extracted again from the LR image and these image patches were fitted into the sparse coding method called Douglas-Rachford to produce the sparse representation vectors which were then used to compute the denoised image patches. The denoised image patches were averaged and the HR image was also successfully produced at the end of the process. However, in order to choose the best feature which is having the best performance, Table-1 shows the PSNR and SSIM values of the final image produced by using dictionary developed based on each feature mentioned in this paper.

As we can observed from table 1, the denoised image produced by extracting the image contrast feature for dictionary learning process was having a slightly higher result with PSNR and SSIM values of 30.5427 dB and 0.8620 respectively. On the other hand, most of the feature representations were constructing a final image with PSNR and SSIM values of approximately 29 dB to 30dB and 0.8300 to 0.8600 respectively. However, the denoised image obtained by using the learned dictionary based on gradient direction as feature extracted obtained a PSNR and SSIM values of only 27.6676 dB and 0.7881.

The best feature to be used in producing a biometric image which can be avoided from the LR issues is image contrast, while the feature which was having the lowest results on PSNR and SSIM values is gradient direction. Therefore, the gradient direction is not suitable to be extracted as a feature in dictionary learning process as compared with other features.

**Table 1.** PSNR and SSIM values of the final image based on each feature representations

| Features                        | PSNR (dB) | SSIM |
|---------------------------------|-----------|------|
|                                 | Before    | After | Before    | After    |
| Energy                          | 24.4323   | 30.5201 | 0.6053 | 0.8603 |
| Gradient Magnitude              | 24.4449   | 29.3530 | 0.6043 | 0.8341 |
| Gradient Direction              | 24.3995   | 27.6676 | 0.6011 | 0.7881 |
| First Derivative                | 24.4538   | 29.7460 | 0.6046 | 0.8456 |
| Second Derivative               | 24.4375   | 29.6375 | 0.6048 | 0.8455 |
| Standard Deviation              | 24.4257   | 30.4428 | 0.6041 | 0.8583 |
| Maximum                         | 24.4491   | 30.5127 | 0.6034 | 0.8595 |
| Minimum                         | 24.4434   | 29.2636 | 0.6049 | 0.8360 |
| Variance                        | 24.4392   | 30.5013 | 0.6041 | 0.8599 |
| Gradient Mean                   | 24.3986   | 30.0844 | 0.6035 | 0.8504 |
| Standard Deviation of Gradient  | 24.4479   | 30.0638 | 0.6047 | 0.8496 |
| Gradient Variation              | 24.4436   | 29.7206 | 0.6047 | 0.8436 |
| Sum                             | 24.4526   | 30.1282 | 0.6050 | 0.8543 |
| Smoothness                      | 24.4544   | 30.2008 | 0.6052 | 0.8578 |
| Kurtosis                        | 24.4256   | 30.1037 | 0.6036 | 0.8529 |
| Skewness                        | 24.3789   | 30.3521 | 0.6018 | 0.8564 |
| Mean Density                    | 24.4580   | 30.2683 | 0.6048 | 0.8578 |
Table 1. PSNR and SSIM values of the final image based on each feature representations
(Continued…)

| Features                  | PSNR (dB) | SSIM |
|---------------------------|-----------|------|
|                           | Before    | After| Before | After |
| Intensity Variation       | 24.4343   | 30.1216 | 0.6040 | 0.8543 |
| Image Contrast            | 24.4558   | 30.5427 | 0.6054 | 0.8620 |
| Local Entropy Filtering   | 24.4523   | 29.6856 | 0.6049 | 0.8449 |
| Local Range Filtering     | 24.4417   | 29.5667 | 0.6038 | 0.8402 |
| Local Standard Deviation Filtering | 24.4447   | 29.5632 | 0.6044 | 0.8421 |
| Entropy                   | 24.4401   | 30.5342 | 0.6052 | 0.8613 |

5. Conclusion and Future Development
As a conclusion, the results show that when different features were extracted for conducting the dictionary learning and sparse coding algorithm, denoised image with different PSNR and SSIM were produced at the end of the process. After that, the important features and less important feature in generating the final dictionary can also be known from the result obtained. Since gradient direction is producing a final image with the lowest performance in term of PSNR and SSIM values, it is not recommended to be applied in the process of future works. This work can be improved by combination the features for the purpose to get an image with greater resolution in future development. Lastly, the input images can also be increased and improved by using the real face images obtained from either the publicly available or local database.

References
[1] Ayas, S., & Ekinci, M. 2020 Single image super resolution using dictionary learning and sparse coding with multi-scale and multi-directional Gabor feature representation. Information Sciences 512 1264-1278.
[2] Nasrollahi, K. and Moeslund, T. B. 2014 Super-resolution: a comprehensive survey Machine vision and applications 25(6) 1423-1468.
[3] Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H. and Zhang, L. 2016 Image super-resolution: The techniques, applications, and future Signal Processing 128 389-408.
[4] Banmore, V. 2009 Iterative-interpolation super-resolution image reconstruction Studies in Computational Intelligence 195 1-8.
[5] Ndikumana, A., Tran, N. H., Ho, T. M., Niyato, D., Han, Z., and Hong, C. S. 2018 Joint incentive mechanism for paid content caching and price based cache replacement policy in named data networking IEEE Access 6 33702-33717.
[6] Jiang, C., Zhang, Q., Fan, R., and Hu, Z. 2018 Super-resolution ct image reconstruction based on dictionary learning and sparse representation Scientific reports 8(1) 1-10.
[7] Umehara, K., Ota, J., Ishimaru, N., Ohno, S., Okamoto, K., Suzuki, T., ... and Ishida, T. 2017 Super-resolution convolutional neural network for the improvement of the image quality of magnified images in chest radiographs Image Processing 10133 p. 101331P
[8] Umehara, K., Ota, J., and Ishida, T. 2018 Application of super-resolution convolutional neural network for enhancing image resolution in chest CT Journal of digital imaging 31(4) 441-450.
[9] Tian, J., and Ma, K. K. 2011 A survey on super-resolution imaging Signal, Image and Video Processing 5(3) 329-342.
[10] Zhang, Z., Xu, Y., Yang, J., Li, X., and Zhang, D. 2015 A survey of sparse representation: algorithms and applications IEEE access 3 490-530.
[11] Wang, Z., Liu, D., Yang, J., Han, W., and Huang, T. 2015 Deep networks for image super-resolution with sparse prior In Proceedings of the IEEE international conference on computer vision 370-378.
[12] Elad, M., & Aharon, M. 2006 Image denoising via sparse and redundant representations over
learned dictionaries *IEEE Transactions on Image processing* **15**(12) 3736-3745.

[13] Li, S., and Qi, H. 2015 A Douglas–Rachford splitting approach to compressed sensing image recovery using low-rank regularization *IEEE Transactions on Image Processing* **24**(11) 4240-4249.

[14] “Image Databases.” [Online]. Available: http://www.imageprocessingplace.com/root_files_V3/image_databases.htm. [Accessed: 03-Jan-2020].

[15] Ayas, S., & Ekinci, M 2020 Single image super resolution using dictionary learning and sparse coding with multi-scale and multi-directional Gabor feature representation *Information Sciences* **512** 1264-1278.