เทคนิคการสร้างภาพความละเอียดสูงยิ่งโดยใช้การแทนแบบเบาบางด้วยพจนานุกรมเกินสมบูรณ์

นายซีโน พัวโนโม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาตรมหาบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

SUPER-RESOLUTION TECHNIQUE USING SPARSE REPRESENTATION WITH AN OVERCOMPLETE DICTIONARY

Mr. Seno Purnomo

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering Program in Electrical Engineering

Department of Electrical Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2011

Thesis Title	SUPER-RESOLUTION TECHNIQUE USING SPARSE REPRESENTATION WITH AN OVERCOMPLETE DICTIONARY
By	Mr.Seno Purnomo
Field of Study	Electrical Engineering
Thesis Advisor	Assistant Professor Supavadee Aramvith, Ph.D
Thesis Co-advisor	Assistant Professor Suree Pumrin, Ph.D.

Accepted by the Faculty of Engineering, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

..... Dean of the Faculty of Engineering
(Associate Professor Boonsom Lerdhirunwong, Dr.Ing.)

THESIS COMMITTEE

..... Chairman
(Assistant Professor Charnchai Pluempitiwiriyawej, Ph.D.)

..... Thesis Advisor
(Assistant Professor Supavadee Aramvith, Ph.D)

..... Thesis Co-advisor
(Assistant Professor Suree Pumrin, Ph.D.)

..... Examiner
(Assistant Professor Thanarat Chalidabhongse, Ph.D.)

..... External Examiner
(Sanparith Marukatat, Ph.D.)

Thai Abstract

5270848021: MAJOR ELECTRICAL ENGINEERING
KEYWORD: SUPER-RESOLUTION/SPARSE REPRESENTATION/OVERCOMPLETE DICTIONARY/BASIS PURSUIT DENOISING/SPARSE CODING/ELASTIC NET

SENO PURNOMO : SUPER-RESOLUTION TECHNIQUE USING SPARSE REPRESENTATION WITH AN OVERCOMPLETE DICTIONARY. ADVISOR : ASSISTANT PROFESSOR SUPAVADEE ARAMVITH, PH.D, THESIS COADVISOR : ASSISTANT PROFESSOR SUREE PUMRIN, PH.D., 125 pp.

Image super resolution is very important but is also considered as a challenging problem in image applications. The main researches are about how to bring back missing information in generating higher resolution image given there is only information from single low resolution image. Another issue concerns the application of learning based image super-resolution method to real time video. The computational cost poses a limit in generating higher spatial resolution while reducing temporal video resolution.

In this thesis, sparse representation is used as a fundamental method. In training step, we propose an efficient clustering dictionary to design an overcomplete dictionary. Firstly, we prepare training set as an initial dictionary. We then perform efficient sparse coding to generate basis matrix. The error of the dictionary is reduced using singular value decomposition. In solution step, we propose Elastic Net as a solution of sparse representation problem. Experiments demonstrated that our proposed method can generate high resolution images with better visual quality when compared with conventional method such as Bicubic interpolation. Our method can gives smaller Root Mean Square Error (RMSE) value than other known interpolation methods. Smaller RMSE implies the higher accuracy in the recognition of face, license plate, and other objects.

The last issue is how to apply our proposed method in video super-resolution. We proposed combination of learning based and analytical based method to generate high resolution video. The experiments also indicate that our proposed method is more practical than using Bicubic, Lanczos methods, and other learning based video super-resolution, because it can increase spatial resolution without reducing temporal resolution.

Department:	Electrical Engineering	Student's Signature	.....
Field of Study:	Electrical Engineering	Advisor's Signature	.....
Academic Year:	2011	Co-advisor's Signature	.....

Acknowledgements

In the name of Allah, the Most Gracious, the Most Merciful. First praise is to Allah, the Almighty, on whom ultimately we depend for sustenance and guidance.

Firstly, it is my pleasure to thank those who made this thesis possible, my advisor, Asst. Prof. Dr. Supavadee Aramvith, and my co-advisor, Asst. Prof. Dr. Suree Pumrin. Throughout the study, their encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. Thank you for forcing me to explore all of my ability so I can finish my thesis result. Their timely and efficient contribution helped me shape this into its final form. I express my sincerest appreciation for their assistance, patience, and guidance.

I would like to express my appreciation to JICA's project for AUN/SEED-Net, Japan, for supporting financial of my research and study in Chulalongkorn University under collaborative research project titled Video Processing and Transmission.

I also would like to express my appreciation to International School of Engineering, Chulalongkorn University, for conducting my study during these two years.

I owe my deep gratitude to Asst. Prof. Dr. Charnchai Pluempitiwiriyawej, Asst.Prof. Dr.Thanarat Chalidabhongse, and Dr. Sanparith Marukatat as my thesis committee. My gratitude also to Asst. Prof. Dr. Nisachon Tangsangiumvisai, Asst. Prof. Dr. Widhyakorn Asdornwised, Asst. Prof. Dr. Chedsada Chinrungrueng, and Suvit Nakpeerayuth, M.Eng as my lectures.

I am indebted to my colleagues, friends at Video Technology Research Group, Adisorn, Fang, Jip, Ton, Rhaendly, Tien, Simon, Bua, Tae, Ovy, Not, Omar, Tri, Annur, Amir, and other students in Communication division that I cannot mention here. It is a pleasure to pay tribute also to Indonesian Student Association in Thailand (PERMITHA), all of my Indonesian friends in Athens Apartment for your kindly friendship. I also would like to express my appreciation to Indonesian Education Attaché for supporting our activity in Thailand.

Last but very important, I owe my deepest gratitude to my family members. The undying support from them has lifted me over countless barriers. My sincerest thank is to my parents, my queen Fernati, my prince Hudzaifah, my sister and all of my family. I would not be able to complete this thesis without them.

Contents

Chapter Page

List of Tables

List of Figures

Figure Page

CHAPTER I

INTRODUCTION

Video surveillance system is widely used because of the security issues come through this decade. Video surveillance system consists of closed circuit television (CCTV) cameras placed in outdoor or indoor locations where they can monitor activity as it takes place. Video surveillance system is performed to detect, recognize and track certain objects from image sequences, and more generally to understand and to describe object behaviors.

However, majority of CCTV cameras especially for the consumer cameras are the low resolution CCTV cameras. The video from low resolution camera often lacks detail information. The quality of the video also depends largely on the lighting environment. Thus, insufficient image resolution makes it difficult to see the object clearly.

These limitations get worse if video data needs to be used in other applications such as, face recognition, text understanding, and other biometric applications. For face case, if the human face is captured very far from the camera it is difficult to detect and recognize the face. For license plate, low lighting in parking area and the car movement are usually the problems in recognizing the text on the license plate. Subjective justification is still needed to understand the license plate number. Therefore, we need a method to enhance the quality and the resolution of the captured object in the image. Such method is called image super-resolution.

1.1 Literature Reviews

In this thesis, we focus on three research areas, namely, *image and video super-resolution*, *sparse representation solution*, and *dictionary learning for image super-resolution*. As shown in Fig. 1.1, low resolution image is captured by CCTV camera from real world view with down sampling, wrapping, and blurring factor. The purpose of super-resolution is how to reverse the process by learning the correlation between low resolution and high resolution images. This section reviews the relevant research works.

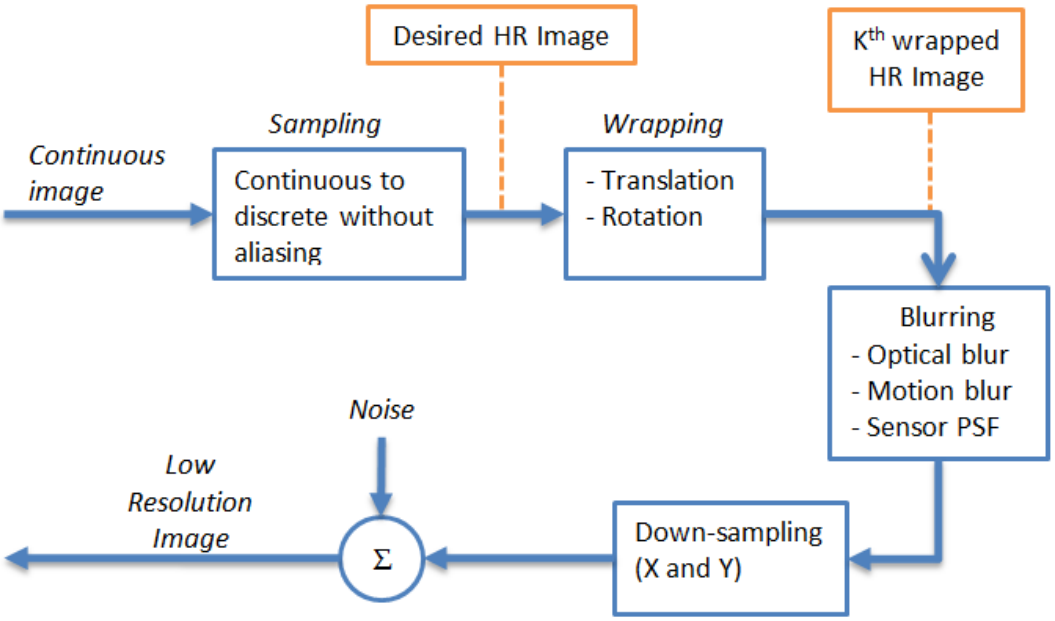


Figure 1.1: Relationship between low resolution image and desired high resolution image.

1.1.1 State of The Art of Image Super-Resolution Techniques

Super-resolution is the way to generate high resolution image from low resolution image. The given low resolution image can be one or more images. Multiple low resolution images can be obtained from image sequences. All of the sources have advantages and disadvantages. By using a single low resolution image, the super-resolution process can be faster and easier. Multiple low resolution images can, of course, provide more information which cannot be found in a single low resolution image.

1.1.1.1 Single Image Interpolation Method

Interpolation or other conventional methods are used to increase the spatial resolution in single image super-resolution. This conventional process is to estimate missing high-resolution detail that is not presented in the original image. Some examples of these conventional methods are categorized based on its purpose. The first one is called smoothing. Its purpose is to make effort to apprehend important information pattern in the image while leaving out noise. Smoothing is one of the interested purpose of Super Resolution to remove noise from the image. The examples of smoothing process are Gaussian, Wiener, moving average, and median filters. Hoever, by using smoothing process, we will loose the image details.

The other purpose is sharpening the image. Sharpening is done not only by increasing the contrast ratio but also by amplifying existing image details. It is useful to increase the sharpness and provided noise is not amplified. But the generated image will give more artifacts. Another method is pixel or block interpolation such as nearest neighbor, bilinear, bicubic, and cubic spline. This methods to combine smoothing and sharpening process. However, this approach cannot provide details in lines, edges, corners, and texture regions.

1.1.1.2 Multiple Image Reconstruction Method

This methods is used when multiple low resolution images captured from the same scene are available. These multiple low resolution images from image sequences sometimes provide different field of view from the same scene. Each low resolution image is naturally shifted with subpixel precision as shown in Fig 1.2. If these images are shifted by integer units, then each image contains the same information, so super-resolution is not possible. But if the low resolution images have different subpixel shifts, then super-resolution is possible. Using this approach, more information can be used to generate the high resolution image.

The first step in this approach is registration. In this step, the estimation of scene motion for each image with reference to one particular image is needed. However, estimating the complete arbitrary motion in real world image scenes is extremely difficult, with almost no guarantees of estimator performance. Incorrect estimates of motion have disastrous implications on overall super-resolution performance.

The second step is to map the pixel information into high resolution pixel grids. Since the shifts between the low resolution images are arbitrary, the images will not always match uniformly to the high resolution image's grid. Thus, non-uniform interpolation is necessary to obtain a uniformly spaced high resolution image from a non-uniformly spaced composite of low resolution images. Non-uniform interpolation between low resolution images is used to improve resolution.

Because blur is usually happened in the interpolation process in previous step, the last step is deblurring. In super-resolution, blur is usually described as a model of a spatial averaging operator.

Dealing with multiple low resolution images to reconstruct high resolution one has some

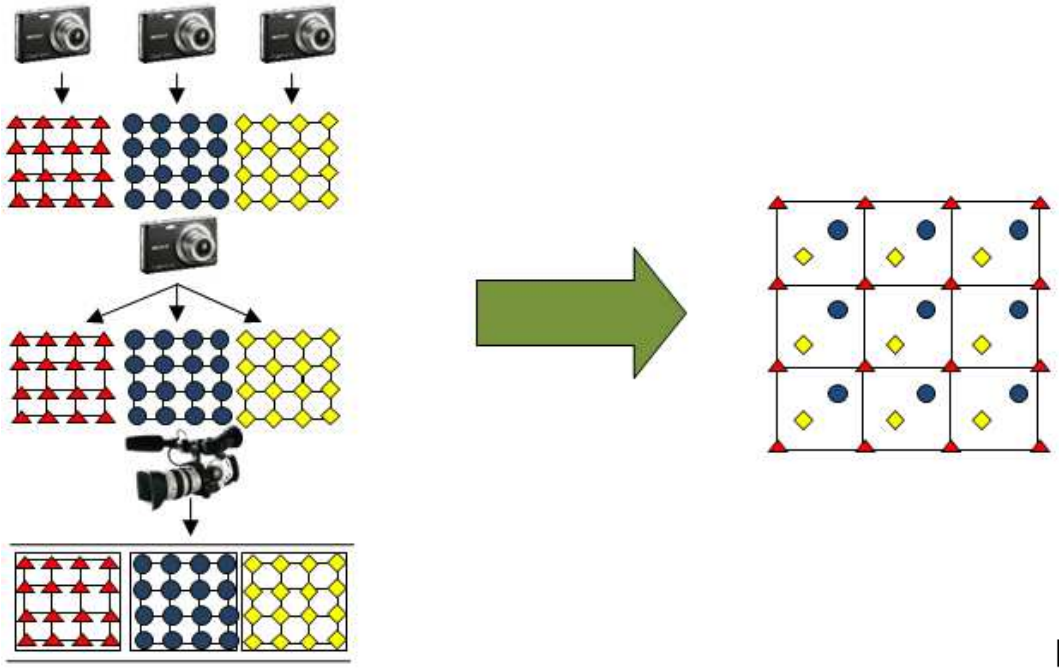


Figure 1.2: Reconstruction of high resolution image given multiple low resolution images.

limitations. This approach does not so efficient when applied to very low-resolution image without consideration of special characteristics of the object. Unfortunately, persons face, license plate, and some body gestures are captured under very low resolution and ambient lighting.

1.1.1.3 Learning Based Method

Because the richness of the real-world images is difficult to capture analytically, many researchers try to find correlation between captured image and real world information. They believe that the correlation information between low and high resolution images can be learned from some training sets. The approach of how to bring back almost all information from high resolution image given very low resolution image is called learning based super-resolution.

Beside of its advantages, learning based approach also has some limitations. The drawbacks of learning based approach always happens on the training process itself. Learning based approach always need the learning dataset. Thus the computational time to learn this dataset of this approach is high. This problem will effect on the time to learn the dataset. The way to optimize this process is still going on and interested to be contemplated now.

Learning based super-resolution tris to find the similarity between image patches in low and high resolution. The similarity is the basic information that stored as the training sets. Generally, this method shares the common idea of using training sets as prior information.

Using only a single image to predict missing high resolution image, learning based method can learn fine detail and estimates missing information of the high resolution image.

1.1.2 The concept of Learning based Image Super-Resolution

Learning based image super-resolution can learn fine detail and predict missing high resolution image given single low resolution image by using training set as a prior knowledge. This algorithm exploits the correlation between low and high resolution image patches from the training sets and then stores the information into codebook dictionary. In the folowings, several methods of learning base image super-resolution are explained.

1.1.2.1 Nearest Neighbor Based Estimation

The fundamental learning based super-resolution algorithms can be characterized as nearest neighbor (NN)-based estimations. In this algorithm, during the training phase, pairs of low-resolution and the corresponding high-resolution image patches (sub-windows of images) are collected. Then, in the super-resolution phase, each patch of the given low-resolution image is compared to the stored low-resolution patches, and the high-resolution patch corresponding to the nearest low-resolution patch is selected as the output.

1.1.2.2 Hallucination

Baker and Kanade [1] proposed training set to generate high resolution face image called face hallucination. This method tries to synthesize a high-resolution face image from an input low-resolution image using gradient prior prediction as a priori term. Gradient descent is chosen to optimize the result. This method still cannot model a prior well and the pixels are predicted individually which may cause discontinuity and noise.

1.1.2.3 Principal Component Analysis

Liu *et. al.* [2] developed a two-step statistical modeling approach that integrates both a global parametric model and a local nonparametric model. First, they derive a global linear model

to learn the relationship between the high-resolution face images and their smoothed and down-sampled lower resolution ones. Second, the residual between an original high-resolution image and the reconstructed high-resolution image using learned linear model, is modeled by a patch-based non-parametric Markov network to capture the high-frequency content of faces.

In face hallucination, the most frequent used subspace method for modeling the human face is Principal Component Analysis (PCA). PCA chooses a new coordinate system such that the variances of the dataset are orderly preserved. Many researchers proposed methods as an improvement algorithm based on PCA.

Wang et. al. applied PCA to the low-resolution face image. In the PCA space, different frequency components are independent. By selecting the number of eigenfaces, we could extract the maximum amount of facial information from the low-resolution face image and remove the noise. The aim of using PCA is to extract as much useful information from low-resolution face image as possible, and then to render a high-resolution face image by eigentransformation. The new hallucinated face image is rendered by mapping between the low and high resolution training pairs.

1.1.2.4 Hallucination by Eigentransformation

Another improvement is face hallucination by eigentransformation, proposed by Wang and Tang [3]. The core idea of the eigentransformation approach is to represent an input LR face image as a linear combination of the LR training samples by using PCA, and then replace the LR training samples with the corresponding HR ones to obtain the estimated HR face. This approach then improved by Hu *et. al.* [4] by using snake algorithm, which is one of the typical curvelet techniques, is employed to automatically segment out the regions of facial features from face images. Then, the eigentransformation is applied to these regions separately. The resulting HR face image is obtained by combining all of the individual reconstructed regions together.

1.1.2.5 PCA and Local Patch Model

Liu *et. al.* [5] proposed a two-step statistical approach integrating the global PCA model and a local patch model. Although the algorithm yields good results, it uses the holistic PCA model to render results like the mean of the image faces. This approach only can represent coarse sharp variation and cannot deal with the detail of high frequency information. It also represents

face poorly since it suffers from overfitting when used for generalization from partial information. Another limitation of holistic PCA method is inability to handle to occlusions. In order to overcome this limitation, local-patch based non-parametric Markov network is applied. But the probabilistic local patch model is very complicated.

1.1.2.6 Non-Negative Matrix Factorization

Yang *et. al.* [6] proposed non-negative matrix factorization (NMF) as a novel approach to the face hallucination problem through sparse coding. This approach was started by the idea that even though faces are objects with lots of variance, they are made up of several relatively independent parts such as eyes, eyebrows, noses, mouths, checks and chins. So Non-negative Matrix Factorization (NMF) was aimed to extract these relevant parts and to find an additive combination of these local features, which is inspired by psychological and physiological principles assuming that humans learn objects in part-based owner. Compared to non-negative matrix factorization, the reconstruction results of PCA are not that intuitive and hard to interpret as PCA allows subtractive combinations of the basis images.

1.1.2.7 Example Based Super-Resolution

All of the previous mentioned methods are focused on enlarging images of a known model class. In real video surveillance application, unidentified object is often found. More general algorithm sometimes is needed to justify what is the object inside the image frame. Freeman *et. al.* [7] proposed basic algorithm in example based image super-resolution. This approach came from the idea that if local image information alone are sufficient to predict the missing high-resolution details; it is possible to use the training set patches by themselves for super-resolution. They modeled the spatial relationships between patches using a one-pass algorithm Markov network. In the one-pass algorithm, they only computed high-resolution patch compatibilities for neighboring high-resolution patches that are already selected, typically the patches above and to the left, in raster-scan order processing.

1.1.2.8 Hybrid Multi-Layer Perceptron - Probabilistic Neural Network

Miravet and Rodrguez [8] proposed two steps neural network based super-resolution. In this approach, after register sub pixel sequences, high-resolution grid nodes are estimated by scattered-point interpolation of the retro-projected image values. This model is used as a training

sample in the neural network. Hybrid Multi-Layer Perceptron - Probabilistic Neural Network architecture (MLP-PNN) is chosen as a core in training process. This image is still degraded by low-resolution optics and detector blurs, as well as by any residual errors of the interpolation procedure. To reduce blur in the image, wiener filter is preferred to use in second process.

1.1.2.9 Combination of Example Based and Conventional Method

Combination of example based and conventional method is successfully proposed by Glasner *et. al.* [9]. In example-based SR, correspondences between low and high resolution image patches are learned from a database of low and high resolution image pairs, and then applied to a new low-resolution image to recover its most likely high-resolution version. Based on the idea that patches in a natural image tend to redundantly recur many times inside the image, both within the same scale, as well as across different scales, combining example based super-resolution with classical method is possible. This method can be applied without using database. Source patch of the block pixel can be learned and founded from the image itself in different scale. Recurrence of patches within the same image scale (at subpixel misalignments) gives rise to the classical super-resolution, whereas recurrence of patches across different scales of the same image gives rise to example-based super-resolution.

1.1.2.10 Sparse Representation

After applied sparse representation in face hallucination, Yang *et. al.* apply their algorithm into general image super-resolution [10]. This method proved that the low-resolution image is viewed as downsampled version of a high-resolution image, whose patches are assumed to have a sparse representation with respect to an over-complete dictionary of prototype signal atoms. This approach use joint learning dictionary pairs between low and high resolution image as a training set. Least absolute shrinkage and selection operator (Lasso) or l1 norm is used as a regularization of linear least square to solve the sparse representation of the predicted high resolution image.

Alder *et. al.* [11] proposed an improvement of single image super-resolution via sparse representation. By exploiting the scale-invariant property of natural images, the set of scalar shrinkage functions are jointly learned from the low-resolution input image. Computer simulations with a simple overcomplete dictionary - the undecimated windowed DCT - revealed superior performance versus the state-of-the-art sparse-representation approach. The proposed Super-Resolution

algorithm includes two stages: during the first stage, a pair of example images is used in an online discriminative learning process of the shrinkage functions. In the second stage, the learned shrinkage functions are applied during the super-resolution reconstruction.

1.1.3 Sparse Representation of Image Super-Resolution

Consider the following regression model with p predictors and n samples:

$$Y = X\beta + \epsilon \tag{1.1}$$

where $X = [x_1, x_2, , x_p]$ are the data, $\beta = [\beta_1, \beta_2, ..., \beta_p]^T$ and $Y = [y_1, y_2, , y_n]^T$ are regressors and response for the ith observation. Let ϵ is the additive noise with dimension $n \times 1$. Suppose the predictors (x_i) are normalized to hare its mean equals to zero and variance equals to one, and the regression output y sums to zero. Ordinary Least Square is a solution to minimize the sum of squared distances between the observed responses in a set of data (Y), and the fitted responses from the regression model (X).

$$\hat{\beta}(OLS) = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||^2 \tag{1.2}$$

The ordinary least square ($\hat{\beta}$OLS) is a technique to estimate the unknown parameters in a linear regression model. The linear least squares computational technique provides simple expressions for the estimated parameters in an OLS analysis, and hence for associated statistical values such as the standard errors of the parameters. It estimates are obtained by minimizing the residual squared error.

1.1.3.1 Ridge Regression

OLS produces a line that minimizes the sum of the squared vertical distances from the line to the observed data points. However, the data analysis with OLS estimates is often not satisfied. The first reason is the prediction accuracy. The OLS estimates often have low bias but large variance; prediction accuracy can sometimes be improved by shrinking or setting some coefficients to zero. By sacrificing a little bias to reduce the variance of the predicted values and hence may im-

prove the overall prediction accuracy. The second limitation is the interpretation. OLS is difficult to interpret a large number of predictors using only very small subsets of observations. A standard technique to improve the OLS estimates is ridge regression, as shown in eq. 1.3. Ridge regression is a continuous process that shrinks coefficients and hence is more stable. This shrinkage has some limitations. It does not give an easily interpretable model because it does not set any unused coefficient to zero.

$$\hat{\beta}(Ridge) = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||^2 + \lambda||\beta||_2^2 \tag{1.3}$$

1.1.3.2 Least Absolute Shrinkage and Selection Operator

To improve the OLS estimate, Tibshirani [12] proposed the least absolute shrinkage and selection operator (Lasso). The proposed method shrinks some coefficients and sets others to zero. The aim is to retain the good features of subset selection and ridge regression. Efron et. al. [13] have provided an efficient and simple algorithm for the Lasso as well as algorithms for stagewise regression. This combination of Lasso and forward stagewise regression is called least angle regression or LARS. LARS algorithm (Efron et al) provides a way to compute the entire lasso coefficient path efficiently at the cost of a full least-square fit. Lasso as a shrinkage of OLS is shown in eq.1.4.

$$\hat{\beta}(Lasso) = \underset{\beta}{\operatorname{argmin}} ||y - X\beta||^2 + \lambda||\beta||_1 \tag{1.4}$$

Let n be the number of variables and p be the number of observations. In case of $p > n$, the lasso selects at most n variables before it saturates, because of the nature of the convex optimization problem. Here, Lasso as a penalized regression method to improve OLS and ridge regression, still has some disadvantages. This seems to be a limiting feature for a variable selection method. Moreover, the lasso is not well defined unless the bound on the ℓ_1-norm of the coefficients is smaller than a certain value. Another usual situations is that $n > p$; if there are high correlations between predictors, it has been empirically observed that the prediction performance of the lasso is dominated by ridge regression. If there is a group of variables among which the pairwise correlations are very high then the lasso tends to select only one variable from the group and does not care which one is selected. Thus, when dealing very large variable, Lasso results in low

accuracy.

Another improvement of Lasso is a pathwise-coordinate optimization proposed by Friedman [14]. The purpose of this paper is to explore one-at-a-time coordinate-wise descent algorithms for convex optimization problems with inequality constraints. Coordinate-wise descent algorithms deserve more attention in convex optimization. They are simple and well-suited to large problems. They have found that for the lasso, coordinate-wise descent is very competitive with the LARS algorithm. Coordinate-wise descent algorithms can be applied to problems in which the constraints decouple, and a generalized version of coordinate-wise descent like the one presented here can handle problems in which each parameter is involved in only a limited number of constraints.

1.1.3.3 Elastic Net

Zou et. al. proposed combination of ℓ_1 norm and ℓ_2 norm as shrinkages in OLS estimate [15]. After a location and scale transformation, they assumed that the response Y is centered and the predictor X is standardized. For any fixed non-negative λ_1 and λ_2, they defined naïve elastic net criterion as shown in eq. 1.5.

$$\hat{\beta}(Na\ddot{i}veEN) = \underset{\beta}{\operatorname{argmin}}||y - X\beta||^2 + \lambda_1||\beta||_1 + \lambda_2||\beta||_2^2 \tag{1.5}$$

If $\lambda_2 = 0$, naïve elastic net becomes Lasso and if $\lambda_1 = 0$, naïve elastic net becomes Ridge regression. Naïve elastic net will strictly convex and has unique solution and thus can do group selection. This grouping effect is guaranteed because highly correlated predictors will have similar regression coefficients. Naïve elastic net still has deficiency. Empirical evidence shows that the naïve elastic net does not perform satisfactorily. The reason is that there are two shrinkage procedures (Ridge and LASSO) in it. Double shrinkage introduces unnecessary bias. Re-scaling of Naïve Elastic Net gives better performance, yielding the Elastic Net solution as shown in eq. 1.6. In order to undo shrinkage and for orthogonal design matrix, the LASSO provides an approximately minimax-optimal solution. In order for Elastic Net to achieve the same minimax-optimality, Zou et. al. [15] rescales naïve elastic net by $(1 + \lambda_2)$.

$$\hat{\beta}(ENet) = (1 + \lambda_2)\hat{\beta}_{Na\ddot{i}veEN} \tag{1.6}$$

Explicit optimization criterion for elastic net can be derived as shown in eq. 1.7.

$$\hat{\beta}(ENet) = \underset{\beta}{\operatorname{argmin}} \left(\frac{x^T x + \lambda_2 I}{1 + \lambda_2} \right) \beta - 2y^T X\beta + \lambda_1 |\beta|_1 \tag{1.7}$$

Friedman *et. al.* also improve Lasso, Ridge regression, and Elastic net using cyclical coordinate descent approach. They proposed regularization paths for generalized linear models via coordinate descent [16]. The model is a fast algorithm for estimation of generalized linear models with convex penalties.

1.1.4 Preprocessing Step and Overcomplete Dictionary

The majority of literature on dictionary design can be categorized into two basic approaches: the analytic approach and the learning-based approach. In the analytic approach, a mathematical model of the data is formulated, and an analytic construction is developed to efficiently represent the model. This generally leads to dictionaries that are highly structured and have a fast numerical implementation. This approach is also called as an implicit dictionary, as they are described by their algorithm rather than their explicit matrix. Dictionaries of this type include Wavelets, Curvelets, Contourlets, Shearlets, Complex Wavelet, and Bandelets [17].

The learning based approach suggests using machine learning techniques to infer the dictionary from a set of examples. In this case, the dictionary is typically represented as an explicit matrix, and a training algorithm is employed to adapt the matrix coefficients to the examples. Algorithms of this type include Principal Component Analysis (PCA) and Generalized PCA, the method of optimal directions (MOD), the K-Means-singular value decomposition (K-SVD), and others. Advantages of this approach are the much finer-tuned dictionaries they produce compared to the analytic approaches, and their significantly better performance in applications. However, this comes at the expense of generating an unstructured dictionary, which is more costly to apply. Also, complexity constraints limit the size of the dictionaries that can be trained in this way, and the dimensions of the signals that can be processed [17].

1.1.4.1 Combination of Wavelet and Edgelet

In order to have more flexibility achieve sparse representation, a novel method is proposed by Donoho [18] based on the combination of two image representation methods, the Wavelets and Edgelets together to form an over-complete dictionary. Combination of edgelet and 2-D wavelet in an overcomplete system is aimed to find decomposition that minimize l1 norm of the coefficients to reconstruct the given image.

1.1.4.2 Empirical Risk Minimization

Another approach in designing an overcomplete dictionary for sparse representation is proposed by Horesh and Haber [19]. This approach is based on minimizing the empirical risk, given some training models. The dictionary model is applied for noisy and noiseless dataset. This method can be utilized for a broad range of applications such as: multi-modality, compressive sensing, optimal experimental design and inverse source localization. Computational cost is one problem that needs to be optimized in this framework.

1.1.4.3 K-Means Clustering Dictionary

Another approach is based on K-means clustering, proposed by Liao et. al. [20] They introduced K-LMS algorithm to design an overcomplete dictionary. It generalized the K-Means clustering process, for adapting dictionaries to achieve sparse representation of signals. The difference between K-Least Mean Square (K-LMS) and ordinary K-means is the atoms update. K-mean process applies the average of K samples to update the atoms. In K-LMS, after removing the K components represented by the atom, it computes the residual error and using LMS algorithm to update the atom. Since every time the atom updates along the residual error is decreasing, the dictionary that can be more effective in representing the training samples after several iterations.

1.1.4.4 K-SVD Dictionary

Aharon et. al. [21] proposed dictionary for sparse representation based on K-means clustering process. They introduced new method in sparse representation called K-SVD. K is from K-means and SVD is the Singular Value Decomposition. This dictionary can work for both orthogonal matching pursuit, a greedy algorithm in sparse representation which selects the dictionary atoms sequentially and basis pursuit which suggests a convexification of the problems by

replacing the ℓ_0-norm with an ℓ_1-norm.

The researchers assessed that KSVD has a good prospective. Then several improvements of K-SVD were proposed.

1. Enhanced K-SVD

   Mazhar and Gader [22] introduced Enhanced K-SVD (EK-SVD) algorithm which finds a dictionary of optimized size for a given dataset, without compromising its approximation accuracy. This algorithm tries to discover the correct number of dictionary elements during dictionary learning, for a given dataset. This algorithm tries to learn technique that discovers an optimized number of dictionary elements by reducing redundancies in the learned dictionary. EK-SVD automatically discovers the value of K during the dictionary learning process. EK-SVD uses an approach similar to the Competitive Agglomeration (CA) algorithm to update the dictionary coefficients helps prune seldom used elements. If there are many similar-looking elements, this approach helps retain only those elements that are frequently used or can be used in place of the others. Once the correct number of clusters has been discovered, EK-SVD uses the Matching Pursuits (MP) algorithm to learn a truly sparse and accurate dictionary. The result of the experiments show that a smaller dictionary learned using EK-SVD can achieve the performance of a bigger dictionary learned using the K-SVD algorithm.

2. Subtractive Clustering Dictionary

   The drawback of clustering process in K-SVD can be improved. The clustering process is related to the selection of proper size of the dictionary. Feng et. al. [23] introduced size variable dictionary learning for sub clustering K-SVD (SC K-SVD) in sparse representations. It came from the fact that a relative small dictionary was selected; it might fail to find the sparse linear combination of the given signal. Moreover, a bigger dictionary might introduce redundant atoms, resulting in extra computation burden on sequential processing tasks. Sub clustering K-SVD algorithm, which characterizes its improvement on K-SVD method in two main aspects: (1) an error-driven mechanism is introduced to the dictionary update stage, achieving a better reconstruction result; (2) priority of the atoms guides the refinement of the dictionary. Thus the most important atoms are retained and well refined.

   Subtractive clustering (SC) is a simple and effective clustering method to find cluster centers based on a density measure called. This technique uses the positions of the data points to

calculate the density function, thus reducing the number of calculations significantly. The first cluster center is chosen as the point having the largest potential and then updated. The next cluster center candidate is also selected according to the new potential values. But it is possible to be accepted or rejected as the real center by some rules. SC can be used to improve K-SVD in the following two aspects. Firstly, SC can be used to initialize the dictionary with data points of clustering center. Secondly, we can use SC to pruning similar atoms group or seldom used atoms learned during K-SVD iteration. SC does reduce the dictionary size considerably, but it also has a drawback. When being applied to image patches, it may exclude high frequency atoms compare with those dominant low frequency ones. If we reduce the threshold to retain the high frequency atoms, similar atoms group will also be retained. In order to avoid this situation, it is necessary to category atoms into several groups by their importance and applies SC to each group [23].

There are several differences between EK-SVD and SC K-SVD. The EK-SVD learn dictionaries in the same way as original K-SVD when the dictionary size is set correctly but this algorithm learns more accurate This dictionary is auto sorted by SC K-SVD to extract atom candidates. Another difference is the dictionary size of EK-SVD decreases all the time, but for this algorithm, it increases most of the time with improving approximation quality and decrease only to prune possible similar atoms. As compared with K-SVD method, a rather smaller dictionary is needed to satisfy the given error bound.

3. Sparse K-SVD

Rubinstein et. al. [17] introduced an improvement of K-SVD as a learning-based approach, by combining it with analytical dictionary design approach. They developed Sparse K-SVD, an efficient K-SVD-like algorithm for training the sparse dictionary, and showed that the structure provides better generalization abilities than the non-constrained one. They also state that the generality of the sparse dictionary structure allows it to be easily combined with other dictionary forms. As dictionary design receives increasing attention, the proposed structure of Sparse K-SVD can become a valuable tool for accelerating, regularizing, and enhancing adaptability in future dictionary structures.

Rubinstein et. al. [24] also published their survey results of learning-based dictionaries, analytic dictionaries, and possibility of combination between them.

1.1.5 Super Resolution for Real Time Video

The drawbacks of learning based super-resolution are a large amount of memory requirement to store examples and the high computational cost to find nearest neighbors in the database. Reducing dimensionality of example and the number of samples stored in the database will make the algorithm run faster. But, the minimum size and dimension of sample will not guarantee sufficient information to generate high resolution image. An ideal and efficient data and searching process is needed in video super-resolution.

1.1.5.1 Hardware and Software Improvement

The combination of software and hardware is needed in some algorithm, like the approach which was proposed by Lopez et. al. [25]. A low-cost video super-resolution algorithm together with its implementation onto a general purpose DSP is implemented into chip design to reconstruct high resolution video.

1.1.5.2 Learning Based and Motion Estimation

An approach with dimensionality reduction by DCT and example selection procedure in order to improve the cost of searching procedure and memory requirements is introduced by Watanabe et. al. [26]. Another researcher focused on motion estimation to explore information in order to generate high resolution video [27]. This approach is a hybrid method that combines motion estimation and learning-based super-resolution.

1.1.5.3 Key Frame Based

Other researchers try to explore information in the key frame [26, 28, 29, 30]. Watanabe et. al. [26] use key frame as the training data, and conducts SR for LR frames. The database composes of examples randomly selected in the key frame. They assumed that key frames (used as training data) and super-resolved frames are temporally-closed, they are strongly correlated.

Based on an assumption that a few frames are encoded at normal resolution (key frames) while the other frames are encoded at reduced resolution, Brandi et. al. [28] proposed to use a super-resolution method to up-sample the non-key frames using the key frames as reference. They adopted example-based super-resolution, where they seek to restore the high-frequency informa-

tion of an interpolated block through searching in a database for a similar block, and by adding the high-frequency of the chosen block to the interpolated one.

Brandi et. al. [29] also made an improvement of their algorithm. The proposed method uses motion estimation to find the best-match between a block from the non-key frame and the key frame. To perform the motion estimation, instead of operating on the low-pass versions of the key frame and non-key frame, they applied high-pass filter to the low-pass version and then search for matches. In effect, this band-pass information better predicts the high-frequency bands. After the motion estimation, they have super-resolution output which is a set of best-match high-frequency key blocks to create an estimated high-frequency for the non-key frame.

1.2 Objective of Dissertation

The objectives of this research are

1. Develop an algorithm for adaptive single image super-resolution, for general image and special case for face and license plate.
2. Propose a design of an overcomplete dictionary for single image super-resolution using sparse representation.
3. Extend the developed for super-resolution in video.

1.3 Scope of Dissertation

The scopes of this research in investigating image and video super-resolution are described as follows:

1. Propose sparse representation solution as a method in image super-resolution. In this research, the image super-resolution will be categorized into three groups. The first is general image for all kind of images. The second is the face image super-resolution, which only focuses on frontal face. The last is the text super-resolution which only focused on frontal license plate image.
2. Propose Optimized design of an overcomplete dictionary for single image super-resolution. In this research,the factor such as the effect of the proposed dictionary, the relationship

between dictionary size and error, and the correlation between image resolution and error, are considered in the design of overcomplete dictionary

3. Propose video super-resolution simulation to evaluate the performance and computational cost of the proposed method. In the last part of this research, the performance of the proposed algorithm and dictionary when are applied to video application are shown. The Application will focus on two cases. The first is an enhancement process of face and license plate detection. The second is the super-resolution for real-time video which evaluates the performance and computational cost of the algorithm.

1.4 Synopsis of Dissertation

After the introduction in this chapter, Fundamental concept of learning based super-resolution using sparse representation is explained in **Chapter 2**. In learning based super-resolution, the system can be divided into 2 steps, training step and testing step. Sparse representation of image super-resolution currently becomes the best way to recover information of correlation between low and high resolution image

The algorithm of designing dictionary for learning based image super-resolution is described in **Chapter 3**. This training step is aimed to generate an overcomplete dictionary which contains as many as information of the richness of the world. Several training sets are trained based on different purpose to generate different dictionary. Basic pursuit denoising combine with singular value decomposition is used to train the dictionary. The experiment results indicate that the dictionary has good performance and very effective to generate high resolution image.

Image super-resolution can be applied into two cases, general case and special case. General case means, image super-resolution for general image. There is no classification of the object in the image. Special case of image super-resolution is refered to special object in the image. In **Chapter 4**, special cases which are discussed, are face image and license plate image. In this chapter, Elastic Net as a proposed method gives solution for sparse representation of image super-resolution in powerful performance. Elastic Net can generate high resolution image with lower root mean square error (RMSE). That means, resulted high resolution image is ready to the next step, such as recognition, with higher accuracy.

Learning based system has good performance in recovering missing information in super-

resolution. The limitation of learning based system is on its computational cost. It take longer time than conventional method when generate high resolution image. It will be difficult if learning based method is applied on real-time video super-resolution. In **Chapter 5**, combination of learning based approach and analytic (conventional) method is implemented in video super-resolution. The result is, high resolution video will have good quality in lower computational time.

In order to integrate results from **Chapter 3, 4, 5**, a single user interface has been developed. This interface was built under c# language. This application has some additional features, such as face detection, license plate detection and recognition. Comprehensive report of this application is described in **Chapter 6**

The thesis is finally concluded in **Chapter 7**, together with contribution of each chapter and suggestion for possible future work.

CHAPTER II

FUNDAMENTAL CONCEPT OF LEARNING BASED SUPER RESOLUTION USING SPARSE REPRESENTATION

This chapter describes the basic concept of learning based super resolution using sparse representation. Two main actions in learning based super method are learning process from training set and applying or testing the result from training process. Learning process is the way to design dictionary matrix from a set of training data. Solution step is a procedure on testing the dictionary using proposed rule. Section 2.1 will explain concept of training and testing image super-resolution. General concept of super resolution via sparse representation is described in section 2.2. Our proposed algorithm in super resolution will be mentioned in section 2.3. This chapteris concluded in the last section.

2.1 Basic Concept on Training Process and Testing Procedure of Image Super Resolution

Training and testing procedure are two main steps in learning based image super resolution. Normally, we can take additional step in this process, such as preprocessing before training process and post processing after testing process.

2.1.1 Training Process

Training process is the way to learn what happened when a block of pixels is decimated into smaller resolution. In this process, we have to make an algorithm or procedures to find relationship between high resolution version of the image and the low resolution version.

Since the size of the image is very large, it is difficult to learn the whole imageat once. If the size of the image is 640 x 480 pixels, we have to observe combination of 307,200 pixels in one time. If the image is in true color, we have to observe $3 \times 307{,}200$ variables. It is impossible to design compact and efficient algorithm with this way.

The solution to solve the problem is that we need to partition the image into a number of small block pixels. By using this block pixels, e.g. 5×5, we only need to observe 25 variables of pixel per time. We also can reduce the number of observation by converting RGB image into YCbCr. We only need to observe luminance value of the block pixel.

By dividing image into block pixels and only considering luminance value, we can have a pack of training set. In order to design optimum training process, we have to observe many block pixels. The numbers of block pixels should much larger than the size of block pixels, e.g. 1000 of 5×5 block pixels. To keep information of the whole image, each block pixel should have overlapping area with its neighbor blocks.

To learn correlation between high resolution and low resolution version, we can use a pair of block pixel in one observation. The training process that involves pairs of pairs of low and high resolution image patches is called a joint learning dictionary. Joint learning dictionary construction will solve joint problem between pairs of high and low resolution image patches. The information that can be observed is the feature of the pairs of the image. The features can be luminance, slope, edge, and gradient.

The result of this training process is a matrix that consists of coefficient represent information from each learned block pixel. This matrix can be called dictionary matrix. The dictionary is consist of information from low resolution and high resolution block.

2.1.2 Testing Procedure

. Testing procedure is the way to use the existing dictionary matrix as a reference in super resolution process. The way to perform this procedure has to be match with the way of training process. We have to partition the image into block pixels. Those block pixels are overlapped by one pixel by adjacent block. Then the features of each block are extracted.

Using mathematical equation, we can find the most similar low resolution codebook in the existing dictionary with our block pixel. We use Elastic Net to find correlated basis (patch coefficient) to find reference information in our overcomplete dictionary. Then we can use the information from the dictionary codebook (high resolution information) to generate high resolution version of the block pixel.

After all of the block pixels is generated into high resolution version, the blocks are arranged into high resolution image.

2.2 Sparse Representation for Image Super Resolution

When we take a picture by camera, the captured low resolution image $\boldsymbol{Y}$ is obtained from real view high-resolution image $\boldsymbol{X}$ which is decimated by factor $\boldsymbol{L}$ and adding blur and noise by $\boldsymbol{V}$ filter. We thus can represent the reconstruction constraint, as in Eq. (2.1).

$$Y = LVX \tag{2.1}$$

Sparse representation will helps to present most or all information from a pair of low and high resolution images using linear combination of small number of elements, or called atoms [31]. These atoms are chosen from an over-complete dictionary matrix $\boldsymbol{D}$.

The problem of super resolution is how to recover as much as possible information from degraded high resolution image. In learning based super resolution, it stores information in one matrix, called dictionary. Suppose $\boldsymbol{D} \in \mathbb{R}^{n \times k}$ is an over-complete dictionary that can be represented in sparse representation, as in Eq. (2.2).

$$y_i = D\beta_i \ \ , \ \ i = 1, 2, ..., N \tag{2.2}$$

Where vector $\beta \in \mathbb{R}^k$ contains the representations coefficients of the signal y and $y \in Y$. The error of the sparest solution in Eq. (2.2) represented by P_0, must be less than or equal to ϵ.

$$\min_{\beta} ||\beta||_0 \quad s.t. \quad ||y - D\beta||_2 \leq \epsilon \tag{2.3}$$

Two ways to solve the representation to find β are Matching Pursuit and Basis Pursuit. Matching Pursuit is a greedy algorithm that finds one atom at a time and select atom sequentially to get the sparest solution based on ℓ_0-norm. Basis Pursuit suggests a convexification of the problems mentioned in Eq. (2.2) and Eq. (2.3), by replacing the ℓ_0-norm with an ℓ_1-norm. Basis Pursuit is a convex optimization problem that can be solved via linear programming for decomposing a signals into an optimal superposition of dictionary elements. In words, Basis Pursuit seeks the smallest (in the ℓ_1 norm sense) solution of the underdetermined linear system

$y = D\beta$. Because of its advantages, stability and high estimation accuracy, we use Basis Pursuit to solve super-resolution problem.

Vector $\beta \in \mathbb{R}^k$ as a representative coefficients of the low resolution signal can be found using least square method. Because of the complexity of the signal, constraints are needed when least square is applied. Ridge regression, Least Absolute Shrinkage Selection (Lasso - ℓ_1-norm) [31], or combination between tow constrains will make least square more powerful to find β.

The common problem in image processing is that there are very large number of pixels as variables compared to number of blocks as an observation, i.e. $p \gg n$. With this condition, if we only use one constraint, i.e. Lasso, there are some drawbacks. The drawbacks are the unfulfilled requirement to select at most n variables before it saturates and the inability to do group selection. If there is a group of variables among which the pairwise correlations are very high, then the Lasso tends to arbitrarily select only one variable from the group [15]. Elastic Net as a formula that combines Lasso and Ridge Regression can be used to solve this problem.

$$\beta_{EN} = \underset{\beta}{\operatorname{argmin}} ||y - \beta D||^2 + (1-\alpha)\lambda_2||\beta||_2^2 + \alpha\lambda_1|\beta|_1 \tag{2.4}$$

As shown in (2.4), Elastic Net, as a new method, is powerful to solve this problem, especially to generate high resolution image [32, 33]. Elastic Net has combined ℓ_1 and ℓ_2 penalty where λ_1 and λ_2 are positive weights. If $\lambda_1 = 0$, Elastic Net will be Ridge Regression and if $\lambda_2 = 0$, Elastic Net will be Lasso. Scale factor α is used to avoid double shrinkage to reduce bias, so better performance will be achieved. The elastic net produces a sparse model with good prediction accuracy, while encouraging a grouping effect.

2.3 Proposed Algorithm for Image Super Resolution

In this chapter, overall picture of proposed method is described. The main ideais to build an efficient super resolution based on its sparse representation as shown in Fig. 2.1. The proposed method is matched to basis pursuit and then synchronized to elastic-net as an efficient solution of sparse representation as shown in Algorithm 2.1.

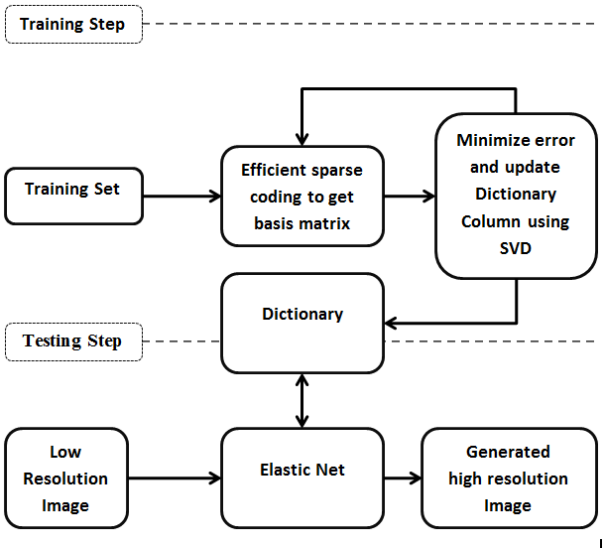


Figure 2.1: Block diagram of the training step and testing step super resolution using proposed method and elastic net.

Algorithm 2.1

Learning Based Image Super Resolution Algorithm

A. Training Step

Repeat until convergence (stopping rule)

1. Initialize Dictionary $D_{i\times j}$
2. Applying efficient sparse coding

   Using Basis Pursuit Denoising solution
2. Dictionary Column Update

   Reducing estimated error using Singular Value Decomposition
3. Save dictionary D

B. Solution (Testing) Step

1. Testing input: a low resolution image Y
2. Upsize Y using Bicubic interpolation
3. For each block pixel y taken starting from the upper left corner with 1 pixel overlap in each direction,

   a. Solve the optimization problem with D and y in (4):

   $\beta_{EN} = \underset{\beta}{\operatorname{argmin}}||y - \beta D||^2 + (1-\alpha)\lambda_2||\beta||_2^2 + \alpha\lambda_1|\beta|_1$

   b. β* = Normalized Coefficients β
4. Output: super-resolution X*

The detail explanation about training step is in chapter 3. Complete explanation of testing step and implementation to several cases can be found in chapter 4. Cases studied in this research are general image, face image, and license plate image.

2.4 Conculsion

In this chapter, the basic concept of learning based super-resolution is described in **Section 2.1**. The idea of sparse representation for image super resolution is explained in **Section 2.2**. In the last section, our proposed method in image super resolution is described in general algorithm. Detail explanation about dictionary learning can be found in **Chapter 3** and implementation of the algorithm in several cases can be found in **Chapter 4**. Performance of the proposed algorithm when be applied in video also can be seen in **Chapter 5**.

CHAPTER III

DESIGN OF AN OVERCOMPLETE DICTIONARY TRAINING FOR LEARNING BASED IMAGE SUPER RESOLUTION

In this section, we discuss about our experiment on designing an overcomplete dictionary for super-resolution using efficient clustering algorithm as an improvement of optimized K-SVD. We also show the performance measurement of the algorithm in term of performance (speed and root mean square error measurement) original K-SVD [21].

3.1 General Concept of Dictionary Training Step

Training step is the way to design an overcomplete dictionary for image super-resolution via sparse representation. Procedures in this training step are started from preparing image dataset to be registered as a training set. The next step is to perform the proposed effective dictionary method to the training set to get an overcomplete and effective dictionary.

3.1.1 Data Set Preparation

Dictionary is a matrix which consists of information needed for super-resolution. The information is obtained from learning process through a training set. In this research, we generate dictionary based on the purpose of super-resolution. Test images include general image, special case for face, and special case for license plate.

For general image, image dataset is taken from Caltech-256 image data set [34]. This image dataset, consist of 256 image categories and each category consist of 30.607 images. Image dataset for face hallucination is taken from Georgia Tech face database [35]. For license plate super-resolution, image dataset is taken from special training set. This special training set should contain information to recover shape and edge of the character on the license plate. Other information such as richness of color can be ignored in this case. So, the training set consists of texture and geometric image dataset.

In this chapter, as a general case of dictionary used in general image super resolution, we use general model of training set from Caltech-256 image data set. Procedure of registering and patch sampling of the training set is presented in Fig 3.1.

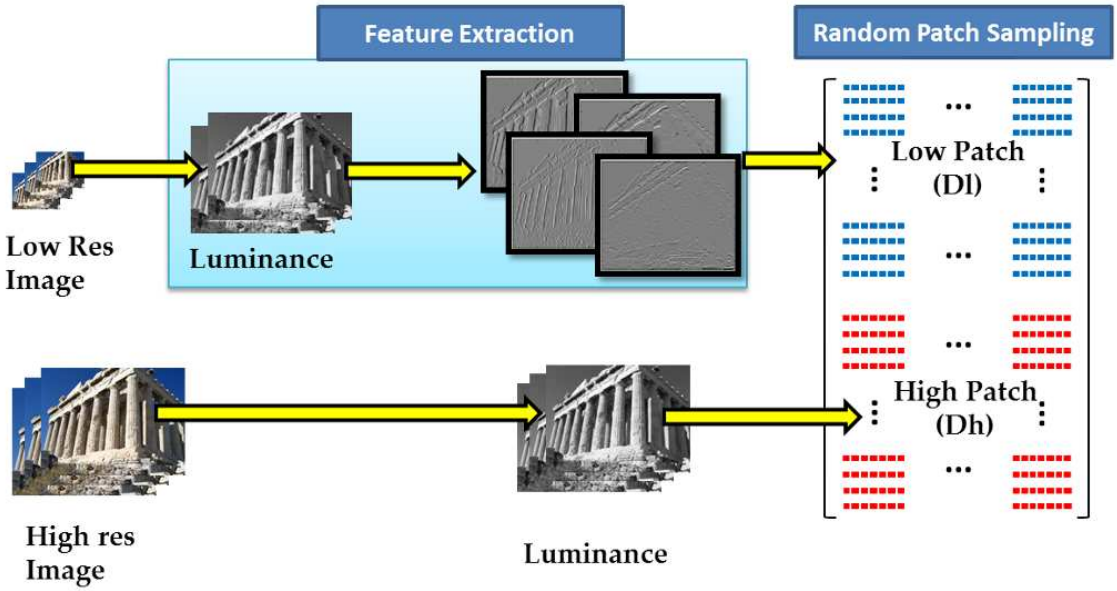


Figure 3.1: Image random patch sampling procedure.

Input of the training process are pairs of high and low resolution images taken from image dataset. Feature extraction is performed to low resolution input. We extracted the feature of low resolution image by taking first and second order gradient both vertically and horizontally. After that, each image was partitioned into block patches. Several pairs of block patches from low and high resolution images are taken randomly. Collected pairs of block patches are registered as a training set. This training set is ready for joint learning dictionary process.

3.1.2 Dictionary Training Algorithm

In the training step, the process starts from the pairs of low and high resolution image patches. These block patches are converted into column vector. The number of column vector is same as the number of observed block pixels. Matrix, containing low resolution patch information which are stored in column vectors, are denoted as D_l and for high resolution are denoted as D_h. Let m and n are dimension of D_l and i and j are dimension of matrix D_h. D_l's width (m) is the number of column vector (observed block pixels) and D_l's height (n) is the number of variables. Number of variables is obtained from number of pixels in one block, multiply by number of extracted low resolution image. D_h's width (i) is equal to m and Dl's height (j) is equal to number of pixels in high resolution patch.

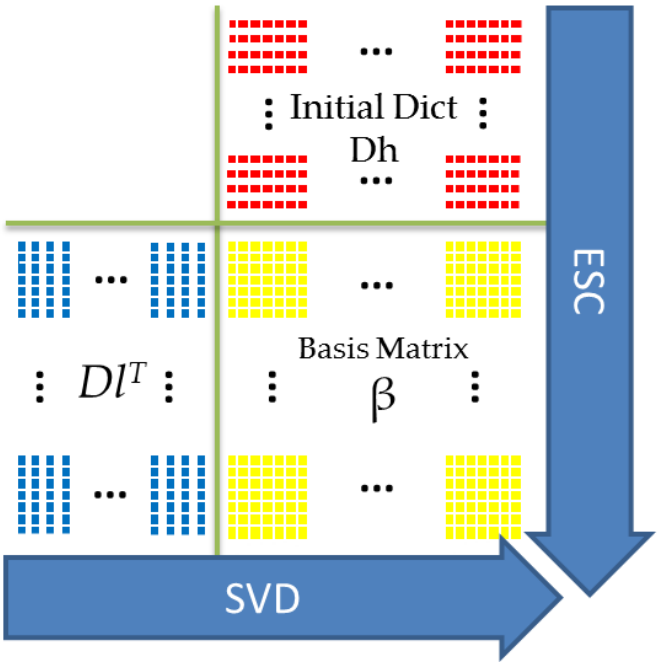


Figure 3.2: Schema of the training process, involving iterative efficient sparse coding (ESC) and singular value decomposition (SVD).

Matrix of D_h is set as an initial dictionary D. Since Dh is the initial dictionary as problem stated in Eq. (3.1), the sparsest basis matrix as a correlation between D_l and D_h can be solved using efficient sparse coding (ESC) of basis pursuit denoising as in Eq. (3.1). In efficient sparse coding, we adopt solution of solving general convex problem of ℓ_1-regularized Least Square using feature-sign search algorithm proposed by Honglak Lee, *et. al.* [36].

Another reason to explain why we use efficient sparse coding is because we expect to use Elastic Net as a solution. Elastic Net is an improvement of Lasso by adding ℓ_2 regularization as a variable selection term. Lasso is known as a function based on Basis Pursuit because it uses ℓ_1-norm as a solution. Because Elastic Net works along basis pursuit, efficient sparse coding of basis pursuit is preferred to use to design the dictionary.

$$||\beta_i||_0 \leq \epsilon \quad s.t. \quad \min_{\beta_i} ||y - D\beta_i||_2^2 \tag{3.1}$$

Like a clustering step of the generalized K-means, dictionary column update is performed. This updating process is aimed to achieve the minimum error E_k between an expected output

with combination of dictionary and basis matrix. The error is the difference between input signal Y and the summation of multiplication of dictionary matrix and basis matrix. Then the error E_k is decomposed using singular value decomposition (SVD) as in Eq.(3.2). Where $d_j \in D$ and $\beta_\epsilon^j \in \beta$. The results of the decomposition are left singular vector U, right singular vector V, and non-zero singular value δ Then matrix U, a result of decomposition, is used to replace column in the dictionary. This process and ESC are proformed iteratively. Schema of this training process is presented in Fig 3.2.

$$E_k = Y \quad - \quad \sum_{j \neq k} d_j \beta_\epsilon^j$$

$$E_k = U \Delta V^T \tag{3.2}$$

The detail algorithm of this proposed dictionary training is shown in Algorithm 3.1. Dictionary generated from this proposed method, then, is evaluated using Elastic Net to generate high resolution image.

Algorithm 3.1

Detail of Dictionary Learning Algorithm in Training Step

1. Initialize Dictionary $D_{h\ i\times j}$
   a. Prepare pairs of high and low resolution image patches as training tet D_h and D_l
   b. Set D_h as an initial Dictionary
   c. D_l as a signal Y is prepared to generate basis matrix (β)
2. Applying efficient sparse coding
   a. Using Basis Pursuit Denoising to approximate minmum value of basis matrix:
      $||\beta_i||_0 \leq \epsilon \quad s.t. \quad \min_{\beta_i} ||y - Dh\beta_i||_2^2$
   b. Do the same on all rows of D_h
3. Dictionary Column Update
   a. Compute error
      E_k=y - $\sum_{j\neq k} dh_j \beta_\epsilon^j$
   b. Apply SVD to get minimum error value
      E_k=$\boldsymbol{U}\Delta V^T$
   c. Do the same on all columns of D_h
4. Repeat 2 and 3 until convergence (reaching minimum E_k) or reaching maximum iteration
3. Save D_h as the Dictionary D

3.2 Experimental Results

The training set for generating dictionary was taken from Caltech-256 image data set [34]. We choose 25 categories (bear, dolphin, duck, fern, grapes, hibiscus, hummingbird, iris, killer-whale, owl, palm tree, penguin, people, raccoon, rainbow, sheet music, skyscraper, snake, stained glass, sunflower, swan, teapot, tennis court, tomato, pisa tower) with 10 images per category. This data set is prepared to be trained in the joint learning process. The resolution of images from the data set is reduced to obtain the low resolution version of the images.

To get as many as information from low resolution image, we extracted the feature of the images. Feature extraction is performed by taking first-order and second-order derivatives using $1-$D filters. First-order vertical gradient filter (f_1) and horizontal gradient filter (f_2) are used to extract information about edge and contour of the image. Second-order vertical derivative filter (f_3) and horizontal derivative filters (f_4) are performed to captures the rate of change in the

intensity gradient.

$$f_1 = [-1\ ,\ 0\ ,\ 1] \quad f_2 = f_1^T,$$
$$f_3 = [-1\ ,\ 0\ ,\ -2\ ,\ 0\ ,\ 1] \quad f_4 = f_3^T \tag{3.3}$$

Then, pairs of patches from low and high resolution images are randomly chosen to obtain two correlated training set matrix D_l and D_h. Subsequently, both D_h and D_l are trained in joint learning dictionary process.

At first, dictionary was trained using original version of KSVD. In this process, the optimization loop is conducted over fifty times. The dictionaries were trained in various sizes, e.g. 256, 512, 768, 1024, 1280, 1536, 1792, and 2048. The proposed dictionary, efficient clustering algorithm, is used to generate the dictionary using three different iterations. The first dictionary clustering optimization was performed in 10 loops, 20 loops, and 30 loops.

Table 3.1: Testing time and RMSE of the generated image using various dictionary and training iteration

Image	Dictionary (# of training iteration)	Test Time (s)	RMSE
Map	KSVD (50 iters)	23.25	6.3489
	Proposed (10 iters)	21.46	6.2911
	Proposed (20 iters)	21.30	**6.2897**
	Proposed (30 iters)	**19.48**	6.3041
Lena	KSVD (50 iters)	7.20	5.6480
	Proposed (10 iters)	7.48	5.5995
	Proposed (20 iters)	6.32	**5.5525**
	Proposed (30 iters)	**6.03**	5.6974
Books	KSVD (50 iters)	17.08	7.9396
	Proposed (10 iters)	7.08	7.8346
	Proposed (20 iters)	8.10	**7.7943**
	Proposed (30 iters)	**6.95**	7.8617
Parthenon	KSVD (50 iters)	4.24	7.0464
	Proposed (10 iters)	3.54	6.8952
	Proposed (20 iters)	**3.29**	6.9203
	Proposed (30 iters)	3.42	**6.8788**
Leaf	KSVD (50 iters)	21.77	**2.8770**
	Proposed (10 iters)	15.75	2.8794
	Proposed (20 iters)	**15.07**	2.8898
	Proposed (30 iters)	18.09	2.8884

3.2.1 Dictionary Performance against Various Pictures

All four dictionaries resulted from the algorithms explained in subsection 3.2 were tested using elastic-net to generate high resolution image from single low resolution image. The testing image was taken from other random images which was not included in the training process. Some of the results of this testing process using 1024-sized dictionary are shown in Table 3.1. The test was performed by 1.37 GHz of Intel Core i7 processor and 4 GB of memory.

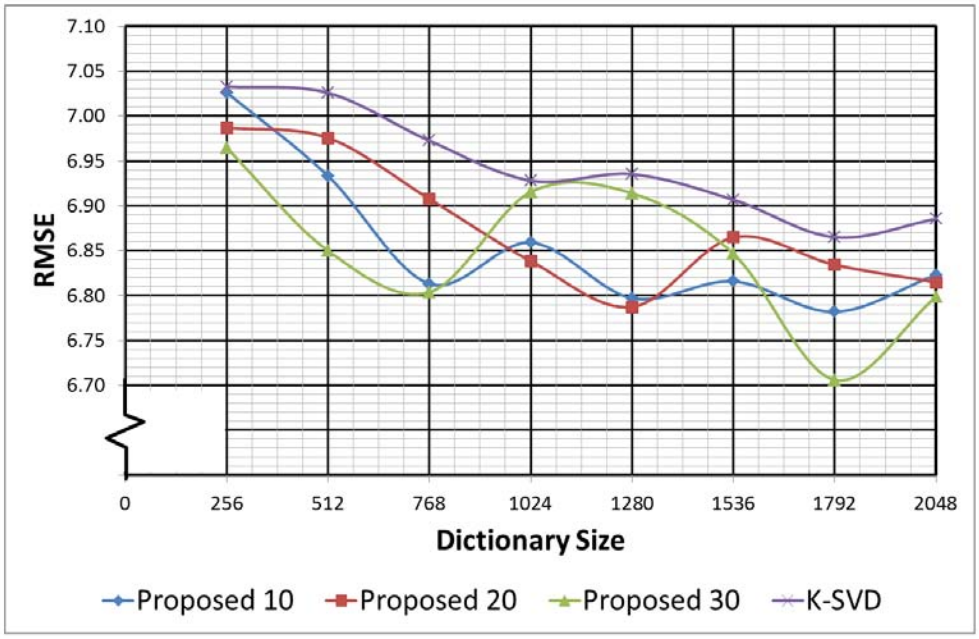
From Table 3.1, it can be seen that the searching time of the elastic-net solution is faster than the proposed dictionary. The dictionary itself was trained in faster iterations (only 10, 20, and 30 times). However, the RMSE of the high resolution image generated by this dictionary was comparable and, in many cases, was lower than KSVD.

Theoretically, if the training steps (efficient sparse coding and dictionary column update) are performed iteratively, residual error value will become smaller. Smaller residual error value means the dictionary is more effective. The performance of dictionary is better as the number of iterations is increased. In summary, the proposed method has advantages in term of fast processing time and comparable quality of the generated high resolution image given the input of a single low resolution image.

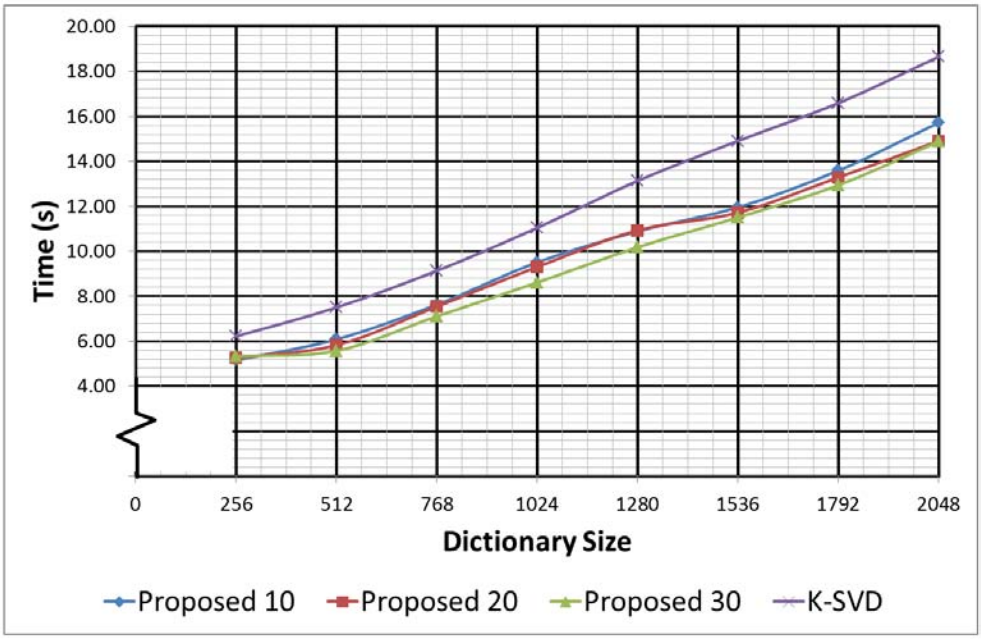
3.2.2 Dictionary Performance against Codebook Size

In order to analyze the performance of the proposed dictionary, various size of the dictionaries has been compared to generate high resolution image. Fig. 3.3 illustrates the results of the RMSE and the time to generate high resolution images using various size of dictionary. The proposed dictionaries were trained in 10, 20, and 30 iterations and then the performance is compared with 50 iterations of KSVD dictionary.

In terms of processing time, the proposed dictionary is faster - gave in the case of 256 to 2048-sized dictionary with 10, 20, and 30 iterations compared to 50 times iterations of KSVD dictionary. The slower time only occur on small size proposed dictionaries with 10 iterations applied on several images. If the dictionary was trained in 20 or 30 iterations, the time to generate high resolution image is usually faster than KSVD trained in 50 iterations. The lower time cost to generate high resolution image makes the proposed dictionary algorithm more preferable to use in image super resolution application. This is because people want to get high resolution image



(a) Dictionary size againts RMSE



(b) Dictionary size againts RMSEr

Figure 3.3: Charts of the average of testing time and RMSE of the generated images using various size of dictionaries and training iterations

faster.

Generated high resolution images form proposed dictionary also have comparable RMSE as KSVD. Moreover, for most of the images, proposed dictionary gives result with lower error than KSVD. This is because proposed algorithm is conducted to store adequate information in better representation.

Size of the dictionary will determine the performance of the system. Larger dictionaries always use more time to generate high resolution image than smaller dictionary. Larger dictionaries almost always give better result with lower error.

3.3 Dictionary Analysis

3.3.1 Analysis of Dictionary Testing Time and Error Measurement

We define time used to test dictionary to generate high resolution image as testing time. Because of the dictionary matrix is fixed, the solvable formula is also fixed (e.g. Elastic Net), thus the error of the generated high resolution image from one same low resolution image is always fixed. Even though the image is tested over and over again, but different things happen in a matter of testing time.

On the testing process, testing time is not fixed for same image. Testing time depends on the speed of the computer and how many applications run at the same time. The testing time also depends on the software used to generate the high resolution image. Nevertheless, the comparison of the time consumption among all of the dictionaries will always be the same.

Different size of the image will also give different testing time. But the complexity of the color and texture of the low resolution image will affect on RMS error much bigger than time consumption. This was evident when the dictionary is used to test the very simple and complex (colored and textured) image as shown in Fig. 3.4 The process to generate the simple image took 10.37 seconds and got RMS error of 0.8084, while for the complex one took 10.71 seconds and got RMS error of 7.4225.

From the calculated testing time of the process and the measured error generate high resolution image, the proposed dictionary gives better performance than KSVD. The proposed dictio-

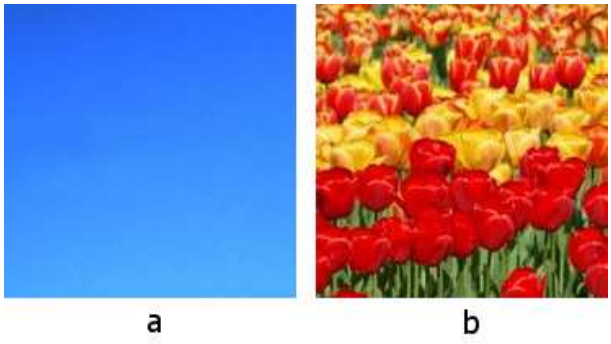


Figure 3.4: Sample of simple image (a) and complex image (b) to show correlation between image complexity and dictionary performance.

nary can be performed faster and gives less error of the generated high resolution images.

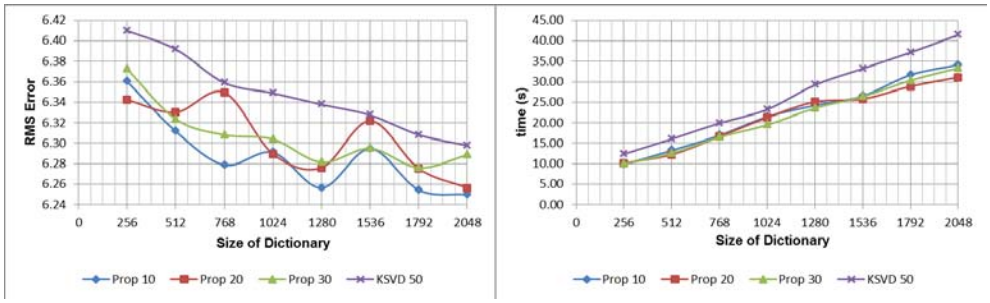
3.3.2 Analysis of Anomaly across Dictionary Performances

The larger dictionaries almost always give better result with lower error. This is theoretically correct, because the larger size of the dictionary means more adequate information are stored. However in some cases, the larger dictionaries can have more error than the smaller one. This incident is called anomaly.

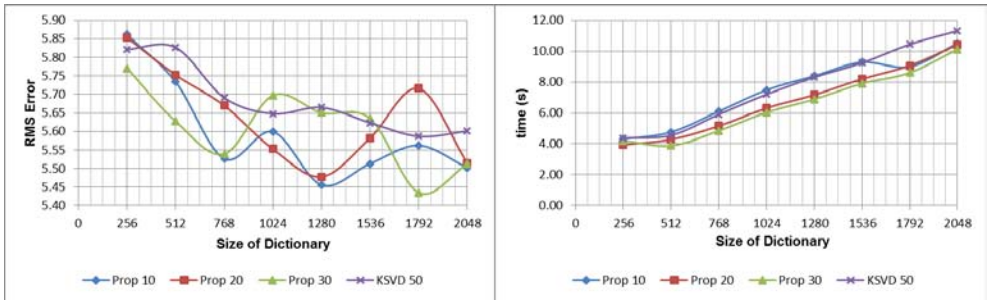
The anomaly can be seen from the graph in Fig. 3.3.b (Lena image) on the proposed dictionary with 20 iterations. When the size of dictionary is 1280, the RMS error is 5.4778, but when the size is increased to be 1792, the RMS error also increases to be 5.7159. In the testing time process, the uncertainties of the result sometimes happen in the unsupervised learning process. Other uncertainty results also can be seen in Table 3.3. Even though our proposed method can give faster results, occasionally higher training iteration cannot give faster testing time.

This anomaly indicates that improper size of the dictionary will give incorrect information. This error could be caused by some ambiguity among two or more observations that are stored in the codebooks. Since the observation data are randomly taken from block pixel in image data set, the degree of difference among all block in the training set cannot be controlled.

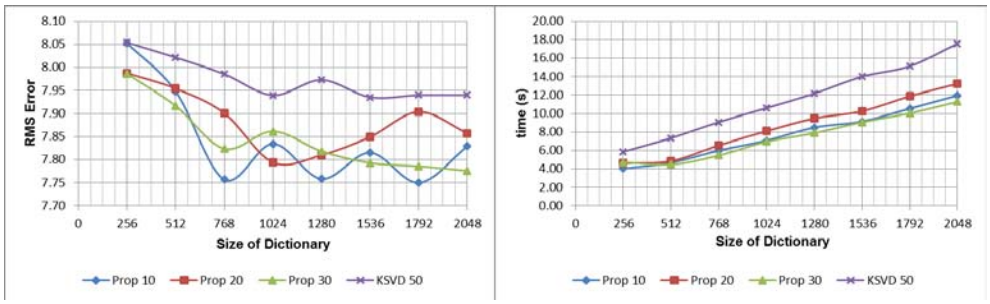
In order to achieve optimum, we have to consider about the richness of the training set, patch sampling process, and how to prune the proposed training set. The other way to reduce the dictionary anomaly, clustering process in the training step should be optimized. The clustering



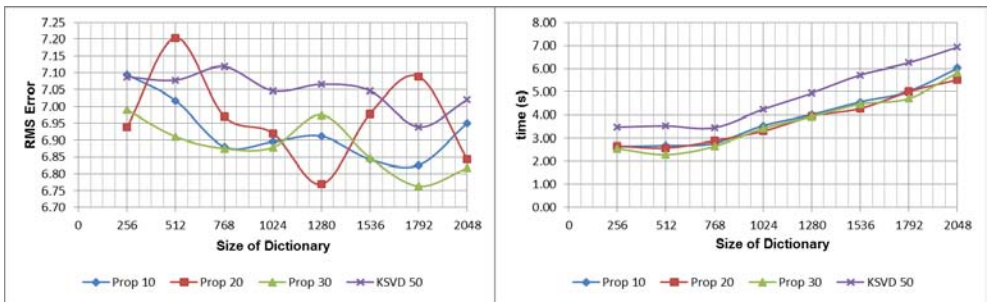
(a) Map



(b) Lena



(c) Books



(d) Parthenon

Figure 3.5: Charts of the testing time and RMSE of the generated image using various dictionaries and training iteration

process is performed to group common observation into the same cluster. In generalized clustering via SVD, each point will belong to a cluster with certain intensity and clusters are not necessarily disjoint. An improvement of SVD applied in very large dimension of matrix is still needed to make the procedure feasible for the very large matrices.

3.4 Summary

We have presented an algorithm to design an overcomplete dictionary for image super-resolution using sparse representation. In super-resolution process, an efficient dictionary is very important. The way to design the dictionary will affect the solution step of the system. The proposed overcomplete dictionary design method, efficient clustering dictionary, gives better performance in term of faster solution speed and lower RMSE. By efficient clustering, the solution can find the best match between tested image and the dictionary codebook faster. The achieved competitive quality result with lower time cost makes the proposed dictionary algorithm more preferable to be used in image super-resolution application. This proposed algorithm is very promising to reduce anomaly performance of the dictionary size and training iterations. Eventually, this dictionary is well suited with Elastic Net solution of sparse representation algorithm to generate high resolution image using super-resolution

CHAPTER IV

IMAGE SUPER RESOLUTION TECHNIQUE USING ELASTIC NET

The resulted dictionary from Chapter 3 gives basic reference on generating high resolution image using its sparse representation. Dictionary learned from proper image dataset will cover much information needed in this process. The proposed dictionary training algorithm gives an overcomplete dictionary matrix that will be used as a reference to generate high resolution image using Elastic Net, as in Eq. (4.1)

$$\beta_{EN} = \underset{\beta}{\operatorname{argmin}} ||y - \beta D||^2 + (1-\alpha)\lambda_2 ||\beta||_2^2 + \alpha\lambda_1 |\beta|_1 \tag{4.1}$$

The algorithm performed in this implementation is shown in Algorithm 4.1

Algorithm 4.1

Learning Based Image Super Resolution for License Plate Recognition Algorithm

---

Solution (Testing) Step

1. Testing input: a low resolution image Y
2. Upsize Y using Bicubic interpolation = Y'
3. For each block pixel y taken starting from the upper left corner with 1 pixel overlap in each direction,
   a. Solve the optimization problem with Dl and y:
      $\beta_{EN} = \underset{\beta}{\operatorname{argmin}} ||y - \beta Dl||^2 + (1-\alpha)\lambda_2 ||\beta||_2^2 + \alpha\lambda_1 |\beta|_1$
   b. $\beta*$ = Normalized Coefficients β
4. Multiply $\beta*$ with Dh as a high resolution patches
5. Use high resolution patches as a correction of Y'
6. Output: super-resolution $X*$

In this chapter, the implementation of Elastic Net as a solution for single image super resolution via sparse representation is discussed. The way to implement Elastic Net in several cases also can be seen in Fig. 4.1. Section 4.1 gives explanation of the implementation on general

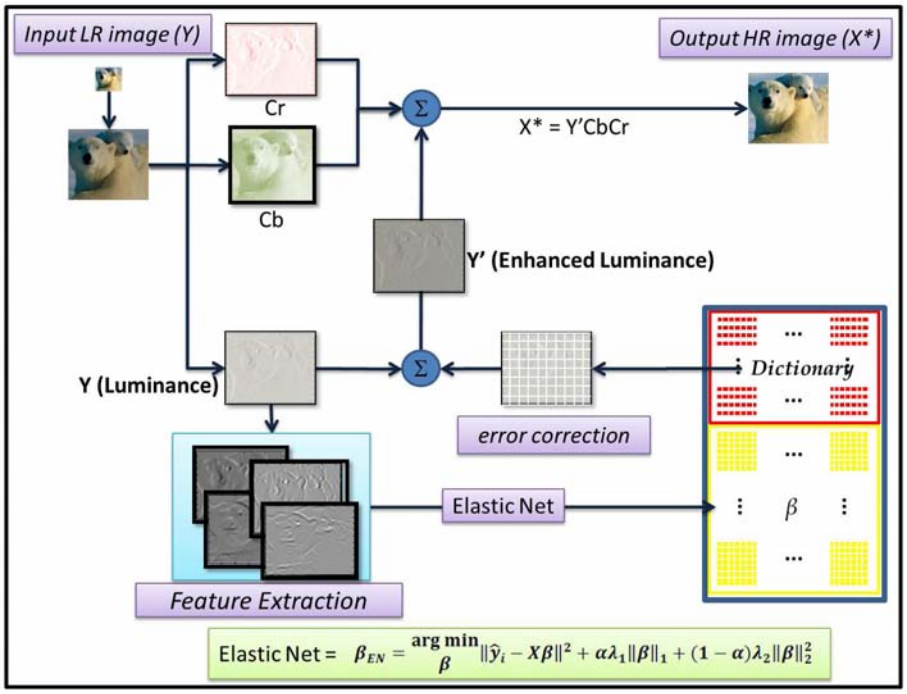


Figure 4.1: Elastic Net is performed to generated high resolution image (X*) from single low resolution image (Y).

image. Analysis of Elastic Net implementation on face hallucination will be explained in section 4.2. The last case of image super resolution that will be discussed in this chapter is license plate case. This case is mentioned in section 4.3. Conclusions is in the last section.

4.1 General Image Super Resolution using Elastic Net

In this section, we will discuss about our experiment applying Elastic Net in super-resolution process. The first case is super-resolution for general image. In general image we do not clasify content of each image. The aim of this implementation is to show performance of our super-resolution method in all kind of images.

4.1.1 Dictionary Design

We generate database using cropped image from Caltech-256 image dataset from California Institute of technology. This dataset consists of 256 image categories containing 30607 images. Complete information about this dataset can be found in [34]. We choose 10 categories (butterfly, iris, fern, galaxy, goldfish, motorbike, mandolin, skycraper, bear, and people) with 20 images per

category.

This database consists of two main parts, D_h for high resolution patches and D_l for Low resolution patches. So we have a pair set of high and low resolution patches $\{X^h, Y^l\}$, where $X^h = \{x_1, x_2, ..., x_n\}$ are the sets of pixels from high resolution image patches and $Y^l = \{y_1, y_2, ..., y_n\}$ are from low resolution image patches.

To try this algorithm, we use various sizes of dictionary. High resolution image is divided into 7 $\times$ 7 pixel patches. Those patches are overlapped by 1 pixel with adjacent patch. It gives size of D_h 49 $\times$ codebook size. Low-resolution image patches size is 3 $\times$ 3 (upsampled to 6 $\times$ 6). We apply first-order and second-order derivatives using $1-$D filters. First-order vertical gradient filter (f_1) and horizontal gradient filter (f_2) are used to extract information about edge and contour of the image. Second-order vertical derivative filter (f_3) and horizontal derivative filters (f_4) are performed to captures the rate of change in the intensity gradient.

$$f_1 = [-1\ ,\ 0\ ,\ 1] \quad f_2 = f_1^T,$$
$$f_3 = [-1\ ,\ 0\ ,\ -2\ ,\ 0\ ,\ 1] \quad f_4 = f_3^T \tag{4.2}$$

Applying those filters gives size of D_l 144 $\times$ codebook size. By joint learning dictionary between D_l and D_h using efficient clustering dictionary described in chapter III, we can generate our overcomplete dictionary for general image super-resolution.

We start from cropping and resizing original images into 180 $\times$ 180 pixels. Those images are high-resolution images. After that, we reduce the size three times to 60 $\times$ 60 pixels. In this section we show some of the results i.e. Parthenon, license plate, kid, and cup. Elastic Net gives better quality in color and texture of high-resolution images as shown in Fig. 4.2.

4.1.2 Experimental Results and Discussions

We also evaluate the effect of the different size of dictionary toward super-resolution performance. We use dictionary which size is 1024. As a comparison, we also create dictionary in a half size (512) and double size (2048). Different size of the dictionary also gives different results. The larger dictionary size, the better high-resolution image is generated by Lasso and Elastic Net.

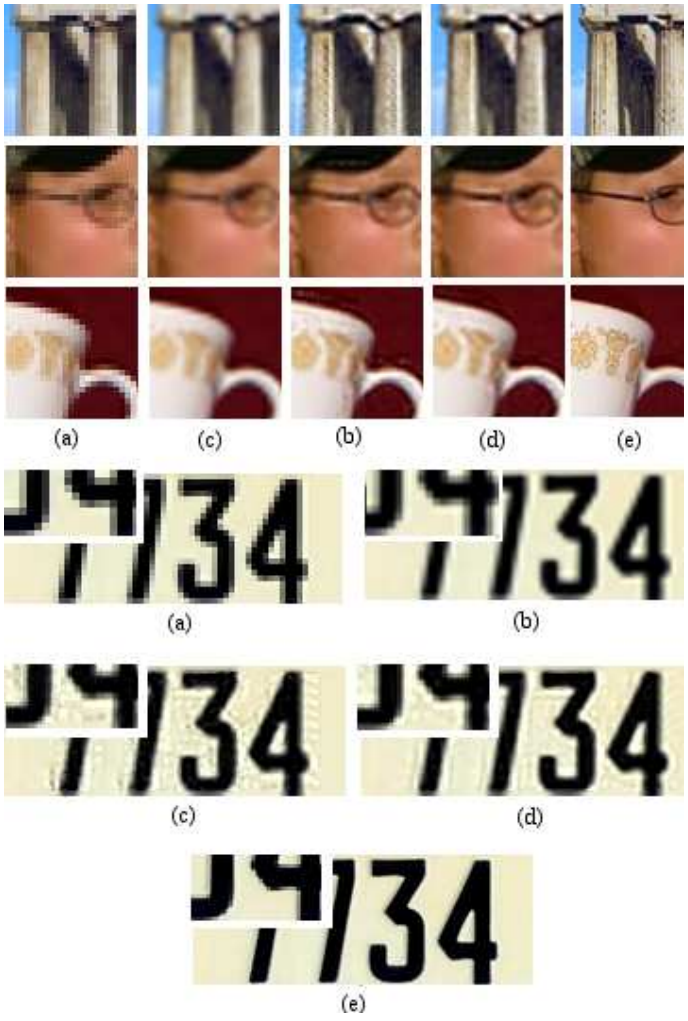


Figure 4.2: High resolution image generated bay our proposed method, (a) low resolution, (b) Bicubic, (c) Lasso, (d) our method, and (e) original image.

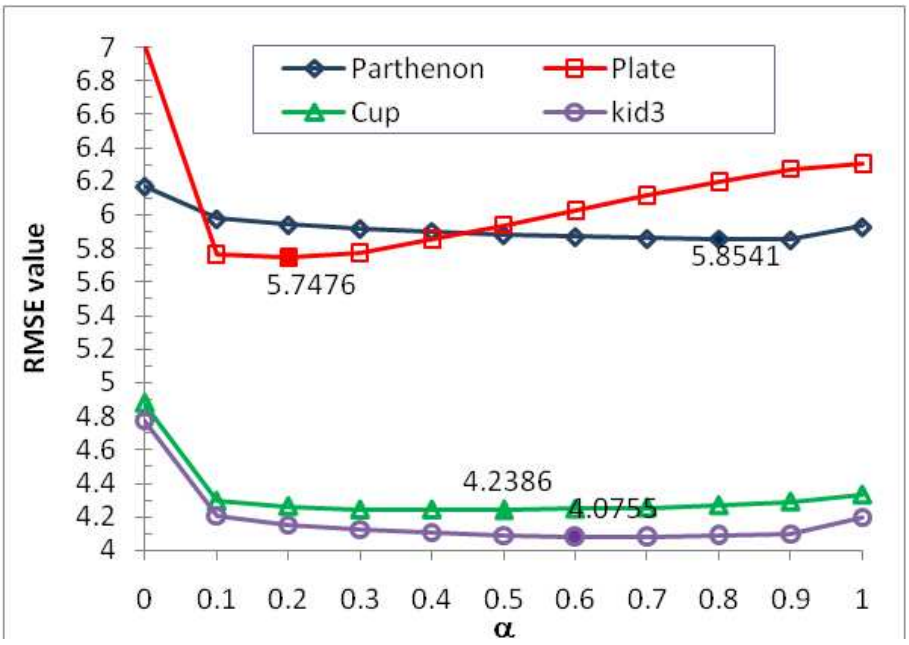


Figure 4.3: The RMS Error of the Generated Image using Elastic Net in Various Value of α (Size of Dictionary = 1024)

As we can see from Table 4.1, Elastic Net gives lower error than Lasso. If the size of dictionary is reduced, the quality of the high-resolution image generated from Elastic Net still smaller than the Lasso one.

We also have to solve the minimum value of β by customizing the value of α. If α = 1, the solution is same as Lasso. And if α = 0, the solution is Ridge Regression. In this part, we want to discuss which value of α will give us the best results with the lowest error. As we see in Fig. 4.3, each image needs different value of α to generate its high-resolution image. Some images need low value of α and the others need high value of α. It means that all kind of images cannot be generalized to solve using Lasso or Ridge Regression. Using Elastic Net, we can compromise

Table 4.1: The RMS Error of the Generated Image using Various Size of Dictionary (with optimum value of α)

Image	**512**		**1024**		**2048**	
	Lasso	EN	Lasso	EN	Lasso	EN
Parthenon	7.5850	7.5401	7.6525	7.4923	7.5580	7.4565
Plate	7.5982	6.8738	7.4267	6.8585	7.1672	6.7803
Kid	4.2554	4.1223	4.2782	4.1479	4.1944	4.1320
Chess	7.4234	7.1286	7.3342	7.1231	7.2763	7.0937

Table 4.2: The RMS Error of the Generated Image using Various Size of Dictionary (with optimum value of α)

Image	**512**		**1024**		**2048**	
	Lasso	EN	Lasso	EN	Lasso	EN
Parthenon	7.5850	7.5401	7.6525	7.4923	7.5580	7.4565
Plate	7.5982	6.8738	7.4267	6.8585	7.1672	6.7803
Kid	4.2554	4.1223	4.2782	4.1479	4.1944	4.1320
Chess	7.4234	7.1286	7.3342	7.1231	7.2763	7.0937

which one is needed, Lasso or Ridge Regression or combination between them.

From all of the experiment results, we know that RMSE value of the high-resolution image generated by our proposed algorithm is lower than previous methods. It means that by using our algorithm, we can recognize an object more accurate than other approaches. This process also take less time than Lasso, about 5 to 25 percent. It means that computational cost of this approach is better than that of Lasso.

4.2 Face Hallucination using Elastic Net

In this section, we will discuss about our experiment applying Elastic Net in super-resolution process. We have applied our proposed method, Elastic Net and Efficient Clustering

4.2.1 Dictionary Design

We generate database using cropped face from Georgia Tech face database. Georgia Tech face database contains images of 50 people from Center for Signal and Image Processing at Georgia Institute of Technology [16]. All of these face images are full faces. All people in the database are represented by 15 color JPEG images with cluttered background taken at resolution 640×480 pixels. The average size of the faces in these images is 150×150 pixels.

We train 210 images which consist of people from many parts of the world (60 European and North American faces, 60 Chinese faces, 30 South Asia (Indian) faces, 15 Latin/South America faces, 30 South East Asian (brown color skin) faces, and 15 African faces). About 90 faces of them are women and the rest are man. From this input training set, about 45 face image wear eyeglasses. The original size of each image is 150 × 150 pixels.

This database consists of two main parts, D_h for high resolution patches and D_l for

Low resolution patches. So we have a pair set of high and low resolution patches $\{X^h, Y^l\}$, where $X^h = \{x_1, x_2, ..., x_n\}$ are the sets of pixels from high resolution image patches and $Y^l = \{y_1, y_2, ..., y_n\}$ are from low resolution image patches.

To try this algorithm, at first, we chose the size of the dictionary to be 1024 patches. As a comparison, we also create dictionary in a half size (512) and double size (2048). High resolution patch size is 7 $\times$ 7 and overlaps 1 pixel with adjacent patch. It gives size of D_h 49 $\times$ codebook size. In order to extract different features for the low resolution image patches, which size is 3 $\times$ 3 (upsampled to 6 $\times$ 6), be applied first-order and second-order derivatives. The four $1-$D filters is used to extract the derivatives are:

$$f_1 = [-1\ ,\ 0\ ,\ 1] \quad f_2 = f_1^T,$$
$$f_3 = [-1\ ,\ 0\ ,\ -2\ ,\ 0\ ,\ 1] \quad f_4 = f_3^T \tag{4.3}$$

Applying those filters give size of D_l 144 $\times$ codebook size. We also apply two zooming factors, three and four times. By joint learning dictionary between D_l and D_h, we can use the same learning strategy in the single dictionary. This joint dictionary is trained using sparse coding algorithm from Honglak Lee *et. al.* [8].

4.2.2 Experimental Results and Discussions

We start from cropping and resizing original faces into 120 pixels. Those images are high-resolution images. After that, we reduce the size three times to 40 $\times$ 40 pixels and four times to 30 $\times$ 30 pixels. Around 45 images have been tasted using Elastic Net. In this chapter, we show the results of 5 different images, ordinary, eyeglasses, mustache, dark skin color, and smiling face. Fig. 4.4 shows that Elastic Net gives better quality in color and texture of high-resolution images.

We also record the RMSE of the high-resolution image in various conditions. Table 4.3 shows that Elastic net gives better quality of high-resolution image than Lasso.

As we show in the algorithm, we have to solve the minimum value of β by customizing the value of α. If α = 1, the solution is same as Lasso. And if α = 0, the solution is Ridge Regression. In this part, we want to discuss which value of α will give us the best results with the lowest error.

Table 4.3: The RMS Error of the Generated Image using various methods (with zooming factor = 4, and optimum value of α)

Image	**Bicubic**	**Lasso**	**Elastic Net**
Ordinary	3.3030	3.6165	**3.2389**
Eye Glasses	5.6413	5.7624	**5.5193**
Mustache	5.4153	5.6894	**5.3853**
Dark	2.9700	3.0988	**2.9686**
Smling face	4.4445	4.4560	**4.1624**

Table 4.4: The RMS Error of the Generated Image using Various Size of Dictionary (Zooming Factor =3, α=0.75)

Image	**512**		**1024**		**2048**	
	Lasso	EN	Lasso	EN	Lasso	EN
Ordinary	3.2911	2.9038	2.8908	2.7484	2.8507	2.7325
Eye Glasses	5.3519	5.0635	5.0901	5.0292	5.0639	4.9849
Mustache	5.1693	4.9230	4.8156	4.7236	4.7802	4.6963
Dark	2.6652	2.4774	2.5013	2.3899	2.4849	2.3789
Smiling	3.9071	3.5336	3.5057	3.4096	3.4559	3.3472

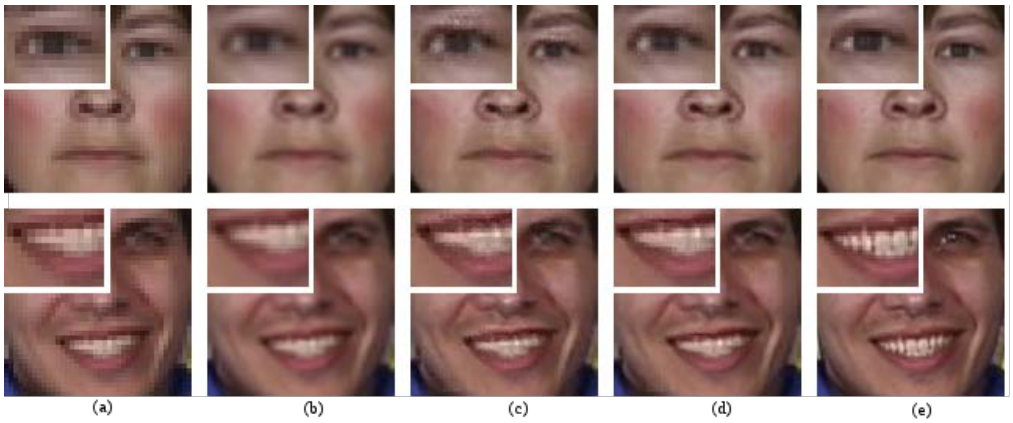


Figure 4.4: High-resolution image generated by Lasso and Elastic Net in various zooming factor: (a) Lasso zooming factor 4, (b) Elastic Net zooming factor 4, (c) Lasso zooming factor 3, (d) Elastic Net zooming factor 3, (e) original image.

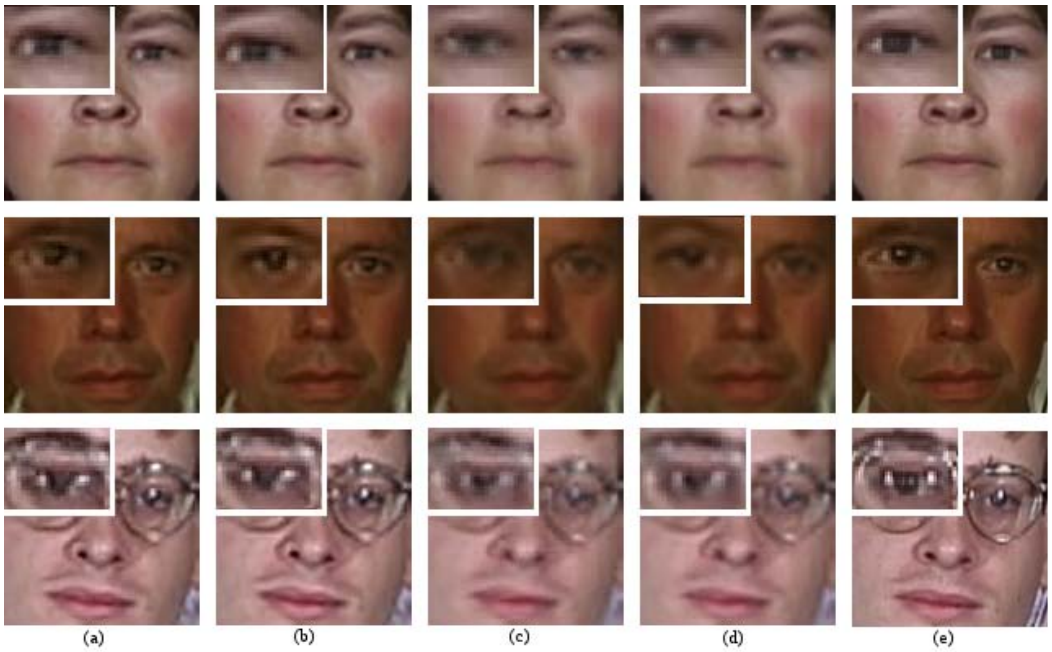


Figure 4.5: High-resolution image generated by various methods: (a) low resolution imge, (b) Bicubic interpolation, (c) sparse representation with Lasso solution, (c) our proposed method, sparse representation with Elastic Net solution, and (e) original high resolution image.

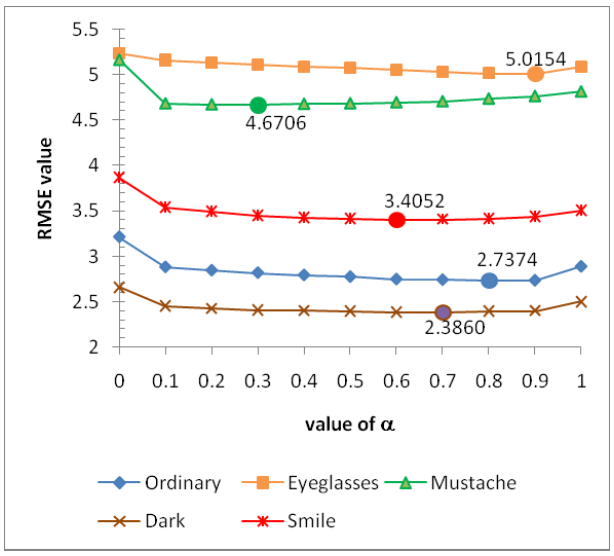


Figure 4.6: RMSE of generated high-resolution image by various value of α.

We also try to make different size of the dictionary to evaluate how far the size will influent the super-resolution process. Different size of the dictionary also gives different results. The larger dictionary size, the better high-resolution image is generated by Lasso and Elastic Net. As we can see from Table 4.4, Elastic Net gives lower error than Lasso. If the size of dictionary is reduced, quality of the high-resolution image from Elastic Net will be smaller than the Lasso one as we can see in Fig. 4.4.

Table 4.5: The RMS Error of the Generated Image using Various (Size of Dictionary = 1024 and optimum value of α)

Image	**3 Times**		**4 Times**	
	Lasso	EN	Lasso	EN
Ordinary	2.8908	2.7374	3.6165	3.2389
Eye Glasses	5.0901	5.0154	5.7624	5.5193
Mustache	4.8156	4.6706	5.6894	5.3853
Dark	2.5013	2.3860	3.0988	2.9686
Smiling	3.5057	3.4052	4.4560	4.1624

To prove the better performance of Elastic Net, we try to enlarge low resolution images four times. At first, a high-resolution image, which size is 120 $\times$ 120 pixels, is down-sampled four times to 30 $\times$ 30 pixels. The low-resolution image is very small; we lose much information during down-sample process. It is more difficult to upsize the resolution and get image clearer. But, using sparse representation and solving with Elastic Net, we still can bring back more information to generate high-resolution image. The results are presented in Table 4.5 and Fig. 4.5.

Table 4.6: The RMS Error of the Generated Image using Elastic Net in Various Value of α (Size of Dictionary = 1024 and Zooming Factor = 3)

Image	α=0	α=0.1	α=0.2	α=0.3	α=0.4
Ordinary	3.2137	2.8846	2.8483	2.8219	2.7966
Eye Glasses	5.2364	5.1630	5.1398	5.1126	5.0892
Mustache	5.1634	4.6844	4.6737	**4.6706**	4.6792
Dark	2.6595	2.4557	2.4268	2.4115	2.4040
Smiling	3.8701	3.5407	3.4941	3.4542	3.4292
	α=0.5	α=0.6	α=0.7	α=0.8	α=0.9
Ordinary	32.7783	2.7512	2.7451	**2.7374**	2.7383
Eye Glasses	5.0800	5.0554	5.0378	5.0207	**5.0154**
Mustache	4.6848	4.6968	4.7079	4.7374	4.7658
Dark	2.3965	2.3873	**2.3860**	2.3978	2.4023
Smiling	3.4154	**3.4052**	3.4107	3.4187	3.4428

As we see in Table 4.6, each image needs different value of α to generate its high-resolution image. Some images need low value of α and the others need high value of α. It means that all images cannot be generalized to solve using Lasso or Ridge Regression. Fig 4.6 also indicated that by using Elastic Net, we can compromise which one is needed, Lasso or Ridge Regression or combination between them.

From the experiment results, we know that RMSE Value of the high-resolution image generated by our proposed algorithm is lower than previous methods. It means that by using our algorithm, we can recognize the face more accurate than other approaches.

4.3 Super Resolution technique on Character Recognition on License Plate

License plate recognition is a process from detecting license plate image on the car until recognizing characters in the image. The problem of license plate recognition is happened when the car is far away from the camera. The dimension of the license plate is too small makes the characters are difficult to be recognized. The other problems come from the error caused by lighting, dust, and car movement. License plate recognition schema is presented in Fig 4.7.

In this chapter, proposed method was applied to upsize detected low resolution license plate image. Using simple optical character recognition (OCR) method, the characters on the license plate ares identified. Then the accuracy is compared with accuracy identification from other conventional super resolution methods using the same OCR. In Fig 4.7, super-resolution will take a part in preprocessing step before the each character is segmented. The proposed method of learning based super resolution for license plate image is shown in Algorithm 4.1.

4.3.1 Dictionary Design

Dictionary for character recognition is generated from special training set. The training set consists of texture and geometric image dataset as shown in Fig. 4.8. The reason of using texture image dataset is to preserve information of plate texture and character's edge and shape. This training set is also aimed to recover the shape to ease the destinction between charcter and background.

The training set is trained using efficient sparse coding of basis pursuit denoising to obtain one dictionary matrix. To reduce the estimated error of the dictionary, singular value decomposi-

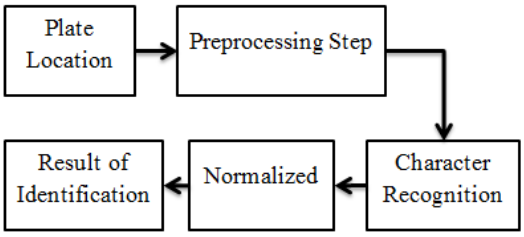


Figure 4.7: Block diagram of character recognition on license plate.

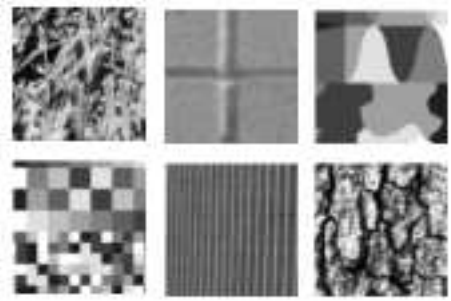


Figure 4.8: Examples of texture image dataset.

tion is applied. This decomposition is performed to update value in each column of the dictionary matrix.

4.3.2 Generating High Resolution License Plate Image

A low resolution license plate image as a part of car was captured from surveillance video. From this image, license plate could be detected using various type of methods, such as simple contour based, otsu thresholding [37], dege density [38], and sobel filter [39].

Detected license plate sometimes is in very low quality and resolution. Using proposed super resolution method, the detected plate is upsized. Compare with conventional method, high resolution license plate image generated from proposed method has better quality, lower noise, and fewer artifacts. This superiority makes the license plate image become better source for character identifiying process.

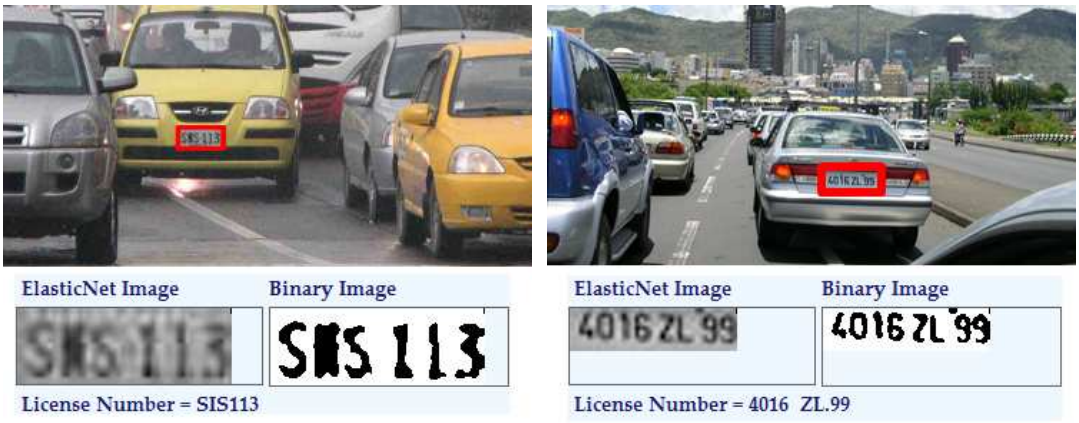


Figure 4.9: Character recognizing on blurry and noisy low resolution license plate images. Low resolution images were upsized using Elastic Net and then converted to binary image. From binary image, license plate number was detected using OCR.

4.3.3 Character Identification from License Plate

The most common way to detect the character is shown in Fig. 4.7. From the detected license plate, we have to separate each character by apply some threshold and then do character segmentation. If generated high resolution license plate image was in low quality, computer will difficult to segment each character. If the character is well segmented and clear enough, the OCR will recognize the character accurately.

We develop an application to simulate super resolution and recognition process under visual C#. This application is used to increase the size of detected license plate image, which is still in low resolution, and then recognize the character on the plate using an open source OCR assembly from Tesseract engine [40].

Based on Fig. 4.9, super resolution was applied on preprocessing step. After low resolution image of license plate was detected, Elastic Net is performed to generate high resolution image. The next step was converting the license plate image into binary image. This step was purposed to distinguish between characters and background and to separate among each character. After separated, OCR would identify each character. If the high resolution image was not in good quality, two or more character will be fused or cannot be differentiated correctly with the background.

Using our proposed method, character information from low resolution image of license

Figure 4.10: Character recognizing from frontal and non-frontal license plate image. Low resolution images are upsized using bicubic and Elasticnet.

plate can be recovered. Fig. 4.9 shows that our proposed method can help OCR to recognize character accurately.

The critical step of detecting characters on license plate is on converting detected plate image to binary image. If object's edge, blur, brightness, and contrast value of the image can be recovered, better quality of binary image can be obtained. That means, OCR can recognize character easily and accurately. As we can see from Fig. 4.10 our proposed super resolution method to increase the detected license plate images is more powerful than conventional method. Our proposed method also can recover almost of all the nearly similar shape of character such as 5 and S; P and R; also B and 8.

Wether, lighting, movement and quality of camera will give noise to the image. Noise and blur make the character blend with the background. Sometimes the plate become darker than the car or the frame and sometimes lighter. This incident makes recognizing process more difficult. Results on Fig. 4.11 also show that our proposed super resolution method can deal with blur and noise from detected low resolution license plate images.

Figure 4.11: Character recognizing on blurry and noisy license plate images. Low resolution images are upsized using bicubic and Elastic Net.

4.4 Summary

In chapter 4, we have presented the performance of Elastic Net when it was applied in several cases. Elastic Net can give better quality of high resolution image generated by learning based super resolution. Elastic Net also can increase accuracy of surveillance system, since generated high resolution image has lower RMSE. The following are conclusions of each case conducted in this research.

4.4.1 General Image Super Resolution

Since Elastic Net can do some group selection when generating coefficients and compromising between Lasso and Ridge Regression, our method can give better results in solving sparse-representation of super-resolution process than Lasso and other conventional methods. Bigger and more complete dictionary will give better results in generating high-resolution images. By choosing the optimum value of α, Elastic Net will give better results. We need the smallest error in recognizing an object and our method gives smaller RMS Error than other approaches. The future works include the optimization of proposed algorithm and application of super-resolution to other kinds of image.

4.4.2 Face Hallucination

Elastic Net as a natural approach for solving convex problem using l_1 or l_2 norm can give better results in solving sparse-representation of super-resolution process than Lasso and other conventional methods. Bigger and more complete dictionary will give better results in generating high-resolution images. By changing various value of α, it is easier to deal with combination of Ridge Regression and Lasso. We need the smallest error in recognizing suspect′s face, and our method gives smaller RMS Error than other approaches. The future works include the optimization of proposed algorithm and application of super-resolution to other kinds of images.

4.4.3 License Plate Super Resolution

Using our proposed super resolution method, the detected low resolution license plate image can be restored to higher resolution. Missing information from the detected image also can be recovered significantly. Our proposed method also can deal with noise and blur caused by environment, movement, and camera itself. With this powerful method, character recognition results on license plate would be more accurate than using conventional methods.

CHAPTER V

VIDEO SUPER RESOLUTION

Super-resolution for video surveillance is part of video super-resolution enhancement processes. The objective of super-resolution for video surveillance is to add some more additional high frequency information into each single frame. This process is needed since there are several tasks of recognition and detection are in fact based on visual data. In this chapter, information of general concept of proposed super-resolution video surveillance system is provided in section 5.1. Section 5.2 will explain how to design dictionary for video super-resolution system. Analysis of the results are described in section 5.3. The conclusion is in the last section

5.1 Video Super Resolution Concept

Video grabbed from indoor camera and outdoor camera will have different kind of distortion. The environmental noise from both conditions is also different. In outdoor situation, ambient lighting changes throughout the day. The ease of use of CCTV camera is the camera always placed in fixed place. That means, the objects captured from CCTV camera are always similar.

In learning based super-resolution system, the limitation is almost come from computational cost. It always takes considerable amount of time to generate one high resolution frame given only single image. Based on assumption that CCTV camera is always fixed; we proposed combination of learning-based and analytic-based super-resolution to display higher resolution of the video.

Since the video is taken from one fixed area, the correlation between current frame and next frame is very high. That means if we can successfully generated high resolution image form first image, it will be easy to generate next frame. This current frame can be used as a reference for next frame as well as the error correction. Using motion estimation we can find most correlated block between current frame and next frame.

By combining learning based method for initial dictionary and fast numerical method, super resolution for video will be very powerful and fast. Initial dictionary generated by learning based super resolution will provide adequate information to generate high resolution image. Fast numerical method will reduce the size of the dictionary and computational cost.

Figure 5.1: Fixed CCTV camera above junction road.

5.2 Dictionary Design

Super-resolution for video surveillance is assumed as increasing resolution and quality of each frame in the video. Each frame will be treated as general image super resolution in Chapter 4. The problem is, it will take many time to generate each single high resolution frame. It is caused by the size of dictionary (1024).

Algorithm 5.1

Initial Dictionary for 1^{st} Frame video Super-Resolution

Initial Dictionary using Sparse Representation

1. Set initial dictionary using algorithm 3.1
2. Input: 1st frame video as low resolution image Y
3. Upsize Y using Bicubic interpolation = Y'
4. For each block pixel y taken starting from the upper left corner with 1 pixel overlap in each direction,
   a. Solve the optimization problem with Dl and y:
      $$\beta_{EN} = \underset{\beta}{\operatorname{argmin}} ||y - \beta Dl||^2 + (1-\alpha)\lambda_2||\beta||_2^2 + \alpha\lambda_1|\beta|_1$$
   b. β* = Normalized Coefficients β
5. Multiply β* with Dh as a high resolution patches
6. Use high resolution patches (Err) as a correction of Y'
7. Output: super-resolution X*

Because of the CCTV camera is only capture one area and will not move to other area, the frames in the video is homogeneous. Each frame is very similar throughout all day operation. Based on this fact, the information needed to be added in the first frame is nearly similar to other frames.

Based on all of the considerations, we proposed combination of learning-based and analytic-based dictionary for video super-resolution. Diagrams of this proposed method is shown in Algorithm 5.1.

This first frame is used as a reference frame and the error correction is used as an error reference frame. For 2^{nd} frame and the next frame are processed using analytical computation based on the calculation of the error reference frame. Using block based motion estimation, correlated block pixel between the current frame and the previous frame can be justified. In our experiment we use 3×3 block pixel and overlapped 1 pixel in vertical and horizontal direction. Correlated block pixel index will indicate error correction value which is needed to enhance current block pixel. Then, the error reference is updated to be used for the next frame.

Algorithm 5.2

Error Reference for n^{th} frame video

Error reference for n^{th} frame video

1. Use previous Err as error refference map (E_{Reff}) to current frame
2. use bicubic image from previous frame (Y') as refference frame (F_{Reff})
3. Input: n^{th} frame video as low resolution image Y
4. Upsize Y using Bicubic interpolation = Y'
5. For each block pixel y taken starting from the upper left corner with 1 pixel overlap in each direction,
   a. Find best match block between current frame and RF using block matching algorithm for

   motion estimation
   b. Get correlate error corerction from E_{Reff}
6. Set current error correction Err as next Error reference map (E_{Reff})
7. Add error correction to current Y'
8. Output: Current high resolution image frame $X*$

Figure 5.2: Generated high resolution image from (a) low resolution frame outdoor video using (b)Lanczos3, (c) bicubic, (d) proposed method.

This scenario still leaves problems. The urgent problem is about the reduced frame rate. This is because calculation in motion estimation still needs some times. This will result in some frames that are skipped. Additional scenario is needed to overcome these skipped frames. Suppose frames generated by previous scenario are called keyframe, the skipped frames are inter frames. These skipped frames can be predicted using previous keyframe.

We propose multithreading or multi processing cumputation to generate missing high resolution frames. These frames are generated using bicubic and corrected by current error refference. These frames are used to fill missing highresolution frame between each keyframe.

5.3 Experimental Result of Video Super Resolution

In video super resolution, we started from initial dictionary trained using efficient sparse coding with SVD. This dictionary was used as a reference to find error correction. This error corection was obtained when super-resolution was applied to the first frame. Then, the error correction was mapped into error matrix as a first error map reference. This error map refference was used as a correction of image super resolution for second frame. Then, error mapped was

Figure 5.3: Generated high resolution image from (a) low resolution frame indoor video using (b)Lanczos3, (c) bicubic, (d) proposed method.

updated according to the information from second frame. This error map would be used as a reference for the third frame and then the error map was updated again. The process is iterated until the last frame of the video.

If there are some significant changes in the environment covered by CCTV camera, this super-resolution process can be refresed. The refreshing process means generating high resolution image frame using Elastic Net with reference from the dictionary like in initialization process. In refreshing time, the current frame input is treated as the first frame. This refreshing process is very useful to keep accuracy and quality of the video.

The way to find the best correlated error correction was to use simple motion estimation from the current frame to the next frame. The result is the video can be upsized faster than using only learning based method. However, the quality is better than using only the conventional method. We compare our method with Bicubic interpolation and Lanczos interpolation. The results indicate the quality of our proposed method is as good as the learning based method but the speed is comparable with the conventional interpolation calculation. Generated high resolution keyframes are shown in Fig. 5.2 and Fig. 5.3.

Figure 5.4: Generated high resolution frame from proposed video super-resolution method.

Problems of outdoor surveillance video super-resolution come from motion blur, lower ambient lighting, and environmental noise. As we can see in Fig. 5.2, our proposed method is more robust againts these problems compared with other conventional methods.

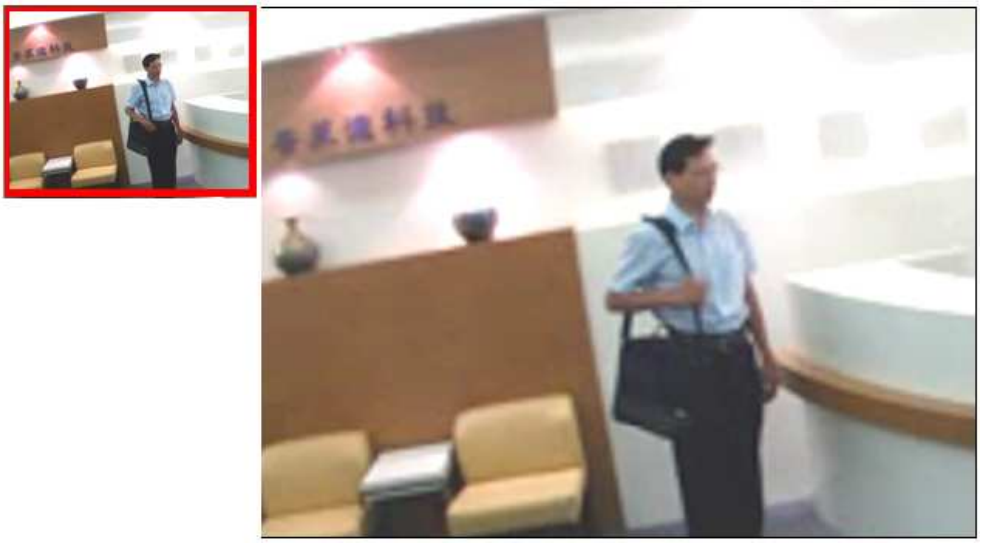
Eventhough most of indoor video surveillance have adequate lighting, enhancement process is still needed. Enhancement process in video surveillance aims to increase accuracy of object recognition application. As seen in Fig. 5.3, our proposed method can improve quality of the face image captured by CCTV camera.This improvement indicates that the accuracy of face recognition is improved.

The resulted image frames for video super-resolution are presented in Fig 5.4.

5.4 Summary

We have proposed method to generate high resolution video frame. Combination between learning based and analytic method in video super resolution has been proposed. An overcomplete dictionary generated using efficient sparse coding method is used as an initial dictionary. This dictionary then becomes error correction map which have compact and with suitable size. This error correction map is updated in each frame and is used as an error correction reference for the next frame. For the next frame, this reference frame is updated and synchronized with error

(a) Frame 1



(b) Frame 2

Figure 5.5: Two sample frames from generated high resolution video using proposed method.

reference map. This combination method gives faster solution in generating high resolution frame without reducing quality of the result.

CHAPTER VI

SUPER RESOLUTION APPLICATION DEVELOPMENT

In order to compile all of the applications in our research, we built super-resolution software. This software was built using Microsoft Visual c#. In this software, we simulated super-resolution for general image, face, license plate, and video. Comprehensive review of this software is described in section 6.1. Summary result is mentioned in the last of this chapter.

6.1 Super Resolution Application Interface

Super-resolution application interface was developed to simulate performance of our proposed method. This interface is based on three case of super-resolution as in chapter 4 and video super-resolution from chapter 5. In this section, we will show analysis of problem in designing interface for super-resolution application using Microsoft Visual C# and the screen shots of the application.

6.1.1 Language and Platform

This application is developed using c# language under Microsoft Visual Studio 2010 Ultimate Edition. C# was developed by Microsoft within its .Net 4.0. C# is intended to be a simple, modern, general-purpose, object-oriented programming language. In this application, three main dictionaries are included, i.e. general image, face, and license plate dictionaries. These dictionaries are stored using xml format.

Some libraries are used to help this application work well. Libraries which are used to help computer vision task in this application are EmguCV, AForge, and Accord. Tessnet2 under Tesseract Engine is used as library for optical character recognition. Other libraries, such as Direct Show and Microsoft DirectX, are used to customize the display of general interface.

6.1.2 Issues on Building Super-Resolution Application

Several issues are faced when building this application. The issues that must be solved are face detection, license plate detection, character recognition, and multithreading (multiprocessing) system. The issues are important point to build an integrated super-resolution application.

Face detection is used in face super-resolution or face hallucination part. In order to automatically select face region of the image or video frame, optimized and fast face detection is needed. The algorithm used in this application is based on haar face detection. This application is also featured by multi face detection up to 5 faces per image. To avoid un detected face region, manual detection is also available.

License plate detection and character recognition are needed in license plate super resolution. Simple license plate detection is applied in this application. The way to detect license plate region is based on image contour. The first step is detecting edge using canny edge detection. After that the exact rectangle region is chosen as a candidate for license plate area. To differentiate between rectangle from license plate and other rectangles, some requirements are considered.

These considerations are rectangle's area and ratio between the height and the width. Optical character recognition library is used to recognize character on the license plate. In this application, Tessnet2 a .NET 2.0 as an Open Source OCR assembly using Tesseract engine is chosen to recognize character on the license plate. The problem is, not all characters are supported by this OCR, such as Thai and Arabic, but Chinese, Korean, and Japanese are provided.

The last issue is multithreading system. Multithreading is a system to support program to process multiple thread separately. Multithreading is needed in this application when we want to process more than one task in one time. For example, if we want to generate high resolution image face from camera, we have to capture image from camera, detect face, and do super-resolution in parallel process. If we do these processes in single processor, the next image captured from camera will wait face detection and face super-resolution to be completed first.

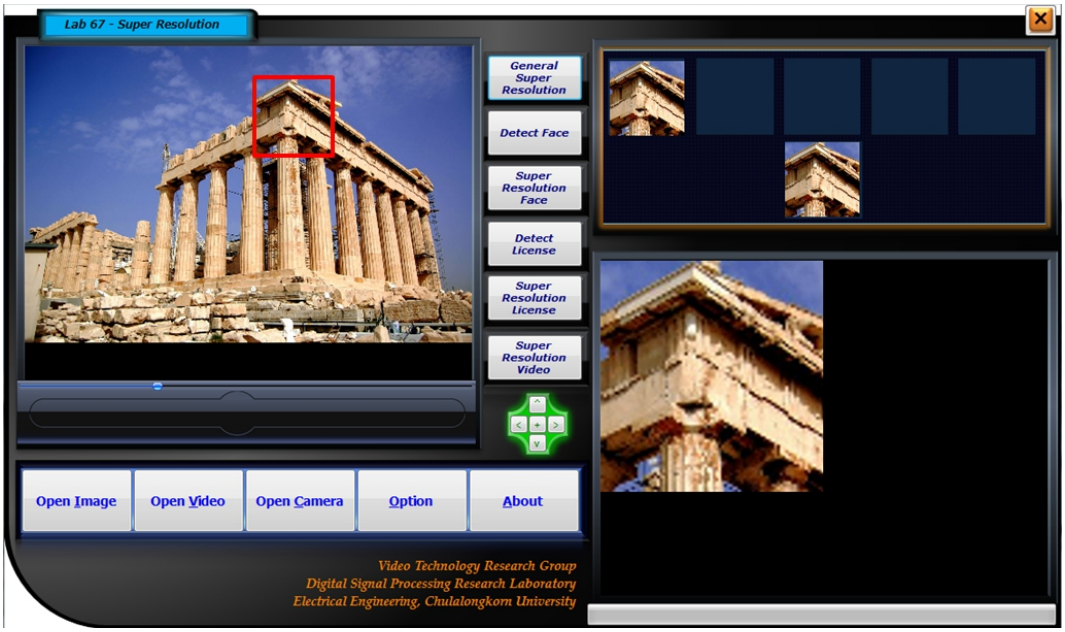


Figure 6.1: Screen shot of general image super-resolution.

6.1.3 Screenshots of Super-Resolution Application

In this subsection main screen shots of the application are captured. These are screenshot for general image super-resolution, face image detection and super-resolution, license plate super-resolution and recognition, and video super-resolution.

This application contains four main applications. Each application will simulate image super resolution from proposed method in **Chapter 4** and video super-resolution from **Chapter 5**. Screen shot of general image super resolution is displayed in fig. 6.1. In this part, user can choose small region of the image and do super-resolution.

Fig. 6.2 is a screen shot for face image super resolution. The source in this part can be image, real time camera capturing, and offline video. User can use both automatic face detection and manual face selection. Maximum numbers of detected faces are 5 images. User can choose which image as a source for super-resolution.

Super resolution interface for license plate recognition is displayed in Fig. 6.3. User can define region of the license plate both manually and automatically. The detected license plate region will be upsized by super-resolution. After high resolution image of the license plate is

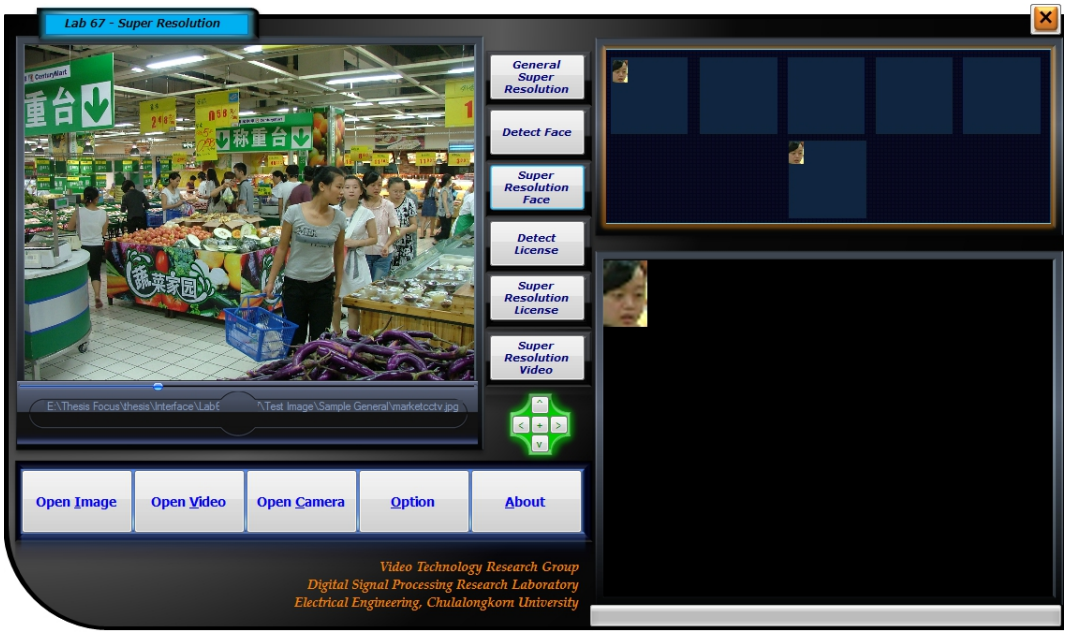


Figure 6.2: Screen shot of face detection and face super-resolution.

Figure 6.3: Screen shot of license plate super-resolution and license plate recognition.

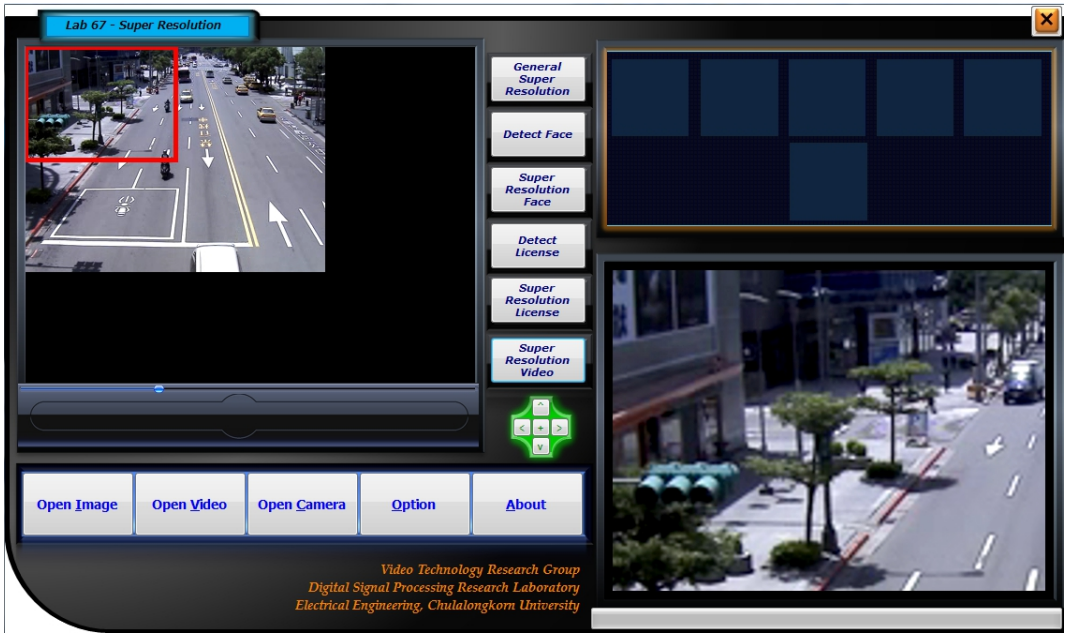


Figure 6.4: Screen shot of video super-resolution.

generated, this application will automatically recognize the characters on the license plate. User can customize how to get binary image from generated high resolution license plate before it be recognized. This customization is embedded in order to increase accuracy of the OCR.

In Fig. 6.4, video super-resolution screenshot is shown. The source of video super-resolution can be real-time video from camera or offline video. In this application, the area of the video that will be upsized is 160×120 pixels. That means original image frames of the video are cropped into the size. User can choose the area of the video that he need to upsized.

6.2 Summary

We have built an application of super-resolution for video and image using Microsoft Visual C#. The application contains four main application three image super resolution cases and video super-resolution. The source of this application is still image, real time camera capturing, and offline video.

CHAPTER VII

CONCLUSIONS

The aim of this thesis is to propose new design of super-resolution. In this thesis, we have done complete research in learning based image super resolution using its sparse representation. The results of this research are an overcomplete dictionary, effective and fast solution, and application in real time video. Super Resolution Interface also has been developed in this research. Using this application, fast super resolution form many cases can be simulated. Each chapter in this thesis gives contribution to overall super-resolution system. The summary of the contributions of each chapter is shown in the following sections, together with suggestion for possible future work.

7.1 Contributions from Chapter 3

In Chapter 3, the framework in designing an optimized dictionary training for super-resolution via sparse representation. This proposed training algorithm is aimed to design an overcomplete dictionary for super-resolution using sparse representation.

The novelty of this proposed training algorithm are in optimizing clustering process on dictionary training and adjust the dictionary model to the application. The ways to optimize clustering process are by using basis pursuit denoising with feature extraction and then performing singular value decomposition. This concept makes the resulted dictionary match with Elastic Net as a proposed sparse representation solution.

The resulted dictionary performance is better than original clustering dictionary i.e. original-KSVD [21]. Using our proposed dictionary, Elastic net can be performed faster compare with original-KSVD. Even it can solve sparse problem faster, RMSE of generated high resolution image also smaller.

The other contributions of our proposed dictionary algorithm is to apply the dictionary model to the application. An overcomplete dictionary will be obtained from proper training set. In this chapter, we proposed an idea that the training set should be matched with the purpose or

the case of super-resolution. The result of this proposed idea is mentioned in section 4.

7.2 Contributions from Chapter 4

While Chapter 3 provides the way to design an overcomplete dictionary, this chapter provides solution for sparse representation of image super-resolution. Elastic Net as a fast and best solution for sparse representation has many advantages. It can converge faster, has a unique solution, and can do group selection.

The contributions in this chapter are to apply Elastic Net with shrinkage scale factor in image super resolution and show the performance in some cases of super resolution. The cases studied in this chapter are general image super-resolution and special image super-resolution. For special case, Elastic Net is performed to generate high resolution face image and license plate image given single low resolution image.

The purpose of the research in this chapter is how to increase accuracy of video surveillance system. By using proposed face hallucination algorithm, face recognition can be more accurate. The proposed algorithm also can be applied on license plate image. Generated high resolution license plate image using our proposed method can help OCR system recognize character on the license plate accurately.

From some experiences conducted in this research, Elastic Net can solve sparse representation effectively. Using Elastic Net, image super-resolution can generate better quality of high resolution image. The quality is measured by RMSE of generated high resolution image using Elastic Net compare with original high resolution image. Lower RMSE as an indication of higher quality can be achieved. It also means, higher accuracy on surveillance application.

7.3 Contributions from Chapter 5

Combination between learning based and analytic method in video super resolution has been performed. An overcomplete dictionary generated using efficient sparse coding method is used as an initial dictionary. This dictionary then becomes error correction map which have compact but efficient size. This error correction map is updated in each frame and is used as a error correction reference for the next frame.

Since camera used in video surveillance always fixed, Elastic Net is only implemented to generate high resolution of reference frame. For the next frame, this reference frame is updated and synchronized with error reference map. This combination method gives faster solution in generating high resolution frame without reducing quality of the result.

7.4 Contributions from Chapter 6

Application framework to test the performance of proposed method in super-resolution has been developed. This application was built using c# language under Microsoft Visual Studio 2010 Ultimate Edition. This application interface is an implementation of proposed image super-resolution method in chapter 4 and video super-resolution in chapter 5. This application covers image super-resolution for general image, face hallucination, and license plate super resolution. This application also has some additional features. Face detection, manual and automatic, was embedded to help user do face hallucination. Automatic and manual license plate detection was also embedded in this application. The optical character recognition was implemented to recognize number on the license plate automatically. This application interface also covers video super-resolution for real time and offline video.

7.5 Possible Future Works

While the thesis research is self-contained, there are numerous scenarios beyond the scope of this work that could be worth studying. Optimizing calculation is still needed to achieve better performance of super-resolution. Better and compact clustering dictionary can be designed to reduce computational cost without degrading the quality of the results. Currently, mobile based application is very popular. In order to implement learning based super-resolution in mobile device, lower computational cost with higher accuracy is becoming the big issue to be considered. Other applications, such as web based and cross platform application, can be alternative solution in image and video super resolution interface.

References

1. Baker, S. and Kanade, T. Hallucinating faces.. IEEE International Conference on Automatic Face and Gesture Recognition.

2. Liu, C., Shum, H.-Y., and Zhang, C.-S. A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition (2001): 192198.

3. Wang, X. and Tang, X. Hallucinating Face by Eigentransformation. IEEE Trans. Syst., Man, CybernPart C: Applicat. Rev. 35 , 3 (2005): 425434.

4. Hu, Y., Shen, T., and Lam, K. M. Region-based Eigentransformation for Face Image Hallucination. IEEE International Symposium on Circuits and Systems, 2009. ISCAS 2009 (2009): 1421 - 1424.

5. Liu, C., Shum, H.-Y., and Freeman, W. T. Face Hallucination: Theory and Practice. International Journal on Computer Vision 75 , 1 (2007): 115-134,.

6. Yang, J., Tang, H., Ma, Y., and Huang, T. Face Hallucination via Sparse Coding. IEEE International Conference on Image Processing (ICIP).

7. Freeman, W., Jones, T. R., and Paztor, E. C. Example Based Super-Resolution. IEEE Computer Graphics and Applications 22 , 2.

8. Miravet, C. and Rodriguez, F. B. A two-step neural-network based algorithm for fast image super-resolution. Image and Vision Computing 25 (2007): 14491473.

9. Glasner, D., Bagon, S., and Irani, M. Super-Resolution from a Single Image. IEEE 12th International Conference on Computer Vision (2009): 349 - 356.

10. Yang, J., Wright, J., Huang, T., and Ma, Y. Image Super-Resolution as Sparse Representation of Raw Image Patches. Proc. CVPR.

11. Adler, A., Hel-Or, Y., and Elad, M. A Shrinkage Learning Approach for Single Image Super-Resolution with Overcomplete Representations. The 11th European Conference on Computer Vision (ECCV) (2010): 5-11.

12. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society B 58 (1996): 267-288.

13. Efron, B., Hestie, T., and Tibshirani, R. Least Angle Regression. The Annals of Statistics 32 , 2 (2004): 407-499.

14. Friedman, J., Hastie, T., Hfling, H., and Tibshirani, R. Pathwise Coordinate Optimization. The Annals of Applied Statistics 1 , 2 (2007): 302-332.

15. Zou, H. and Hestie, T. Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society B 67 , 2 (2005): 301-320.

16. Friedman, J., Hestie, T., and Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software 33 , 1.

17. Rubinstein, R., Zibulevsky, M., and Elad, M. Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation. IEEE Trans. on Image Processing 58 , 3 (2010): 1553 - 1564.

18. Donoho, D. L. and Huo, X. Combined Image Representation using Edgelets and Wavelets. Wavelet Applications in Signal and Image Processing VII, in SPIE 3813 (1999): 468-476.

19. Horesh, L. and Haber, E. Overcomplete Dictionary Design by Empirical Risk Minimization. Inverse Problem.

20. Liao, Y., Xiao, Q., Ding, X., and Guo, D. A Novel Dictionary Design Algorithm for Sparse Representations. CSO '09 Proc. of the 2009 International Joint Conference on Computational Sciences and Optimization 1 (2009): 831-834.

21. Aharon, M., Elad, M., and Bruckstein, A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. IEEE Trans. on Signal Processing 54 , 11 (2006): 4311-4322.

22. Mazhar, R. and Gader, P. D. EK-SVD: Optimized Dictionary Design for Sparse Representations. 19th International Conference on Pattern Recognition, 2008. ICPR 2008 (2008): 1-4.

23. Feng, J., Song, L., Yang, X., and Zhang, W. Sub clustering K-SVD: Size variable Dictionary learning for Sparse Representations. 16th IEEE International Conference on Image Processing (ICIP), 2009 (2009): 2149 - 2152.

24. Rubinstein, R., Bruckstein, A. M., and Elad, M. Dictionaries for Sparse Representation Modeling. Proc. of the IEEE 98 , 6 (2010): 1045 - 1057.

25. Lopez, S., Callico, G. M., Member, Tobajas, F., Lopez, J. F., and Sarmiento, R. A Novel Real-Time DSP-Based Video Super-Resolution System. IEEE Trans. on Consumer Electronics 55 , 4.

26. Watanabe, K., Iwai, Y., Haga, T., and Yachida, M. A Fast Algorithm of Video Super-Resolution Using Dimensionality Reduction by DCT and Example Selection. International Conference on Pattern Recognition, 2008.

27. Song, B. C., Jeong, S.-C., and Choi, Y. Video Super-Resolution Algorithm Using Bi-Directional Overlapped Block Motion Compensation and On-the-Fly Dictionary Training. IEEE Trans. on Circuits and Systems for Video Technology.

28. Brandi, F., Queiroz, R. d., and Mukherjee, D. Super-Resolution of Video using Key Frames. IEEE International Symposium on Circuits and Systems, 2008 (2008): 1608 - 1611.

29. Brandi, F., Queiroz, R. d., and Mukherjee, D. Super Resolution of Video Using Key Frames and Motion Estimation. 15th IEEE International Conference on Image Processing, 2008. ICIP 2008 (2008): 321 - 324.

30. Song, B. C., Jeong, S.-C., and Choi, Y. Key frame-based video super-resolution using bi-directional overlapped block motion compensation and trained dictionary. 2nd International Conference on Image Processing Theory Tools and Applications (IPTA), 2010 (2010): 181 - 186.

31. Yang, J., Wright, J., Huang, T., and Ma, Y. Image Super-Resolution via Sparse Representation. IEEE Trans. on Image Processing.

32. Purnomo, S., Aramvith, S., and Pumrin, S. Generalized Image Super-Resolution Technique using Elastic Net. International Symposium on Multimedia Communication Technology (2010): 185-188.

33. Purnomo, S., Aramvith, S., and Pumrin, S. Elastic Net for Solving Sparse Representation of Face Image Super-Resolution. International Symposium on Communications and Information Technologies.

34. Griffin, G., Holub, A., and Prona, P., Caltech-256 Object Category Dataset. tech. rep., California Institute of Technology, 2007.

35. Nevian, A. V. Georgia Tech face database.

36. lee, H., Battle, A., Raina, R., and Ng, A. Y. Efficient Sparse Coding Algorithm. Proc. of Neural Information Processing System (NIPS).

37. Liu, D. and Yu, J. Otsu method and K-means. Ninth International Conference on Hybrid Intelligent Systems.

38. Wu, M.-K., Wei, J.-S., Shih, H.-C., and Ho, C. C. License Plate Detection Based on 2-Level 2D Haar Wavelet Transform and Edge Density Verification. IEEE International Symposium on Industrial Electronics.

39. Mello, C. A. B. and Costa, D. C. A Complete System for Vehicle License Plate Recognition. International Conference on Systems, Signals and Image Processing.

40. Smith, R. Tesseract OCR Engine. Open Source Convention.

Biography

Seno Purnomo was born in 1984 in Jogjakarta, Indonesia. He received the Bachelor of Engineering in Engineering Physic from Gadjah Mada University, Jogjakarta, Indonesia in 2007. He has been pursuing the Master of Engineering degree in Electrical Engineering at Chulalongkorn University, Bangkok, Thailand, since 2009. His research interests include video and image super resolution, video and image processing, and computer vision.

List of Publications

1. ”Design of Efficient Clustering Dictionary for Sparse Representation of General Image Super Resolution”, Accepted in Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2011), October 2011, Xi’an China.
2. ”Super Resolution Technique for Character Recognition on Low Resolution License Plate Image”, appeared in International Symposium in Multimedia and Communication Technology 2011, September 2011, Sapporo, Japan.
3. ”Elastic Net for Solving Sparse Representation of Face Image Super Resolution”, published in International Symposium on Information and Communication Technology 2010, October 2010, Tokyo, Japan.
4. ”Generalized Image Super Resolution Technique Using Elastic Net”, published in International Symposium in Multimedia and Communication Technology 2010, September 2010, Manila, Philippine.
5. ”Super Resolution Technique for Video Surveillance”, published inThe 2012 International Workshop on Advanced Image Technology, 9-10 January, 2012, Ho Chi Minh City, Vietnam (under review)