Elastic Net for Solving Sparse Representation of Face Image Super-Resolution

Seno Purnomo, Supavadee Aramvith, and Suree Pumrin
Department of Electrical Engineering
Chulalongkorn University, Bangkok 10330, Thailand
E-mail: Seno.P@Student.chula.ac.th Tel: +66-8234 09622
E-mail: Supavadee.A@chula.ac.th Tel/Fax: +66-2-218-6909/6912
E-mail: suree.p@chula.ac.th Tel/Fax: +662-218-6484/6488

***Abstract*—Super-resolution is very important in recognizing suspects face in video surveillance system. In this paper, we present an improvement of image super-resolution based on sparse signal representation. The issue of how to deal efficiently with sparse feature has great significance on the quality improvement of generated high resolution image. We propose to use Elastic net to solve sparse representation of image super-resolution process. Elastic Net will compromise between Lasso and Ridge regression to find the best correlated patch between low-resolution and high-resolution image. Experiments demonstrate that using Elastic Net reconstructed high-resolution images have better color and texture quality. It also gives smaller value of Root Mean Square Error (RMSE) than Lasso and other conventional methods. Small RMSE means more accurate when recognizing face.**

***Keywords* – super-resolution, lasso, elastic-net, sparse-coding, ridge regression, sparse representation, surveillance.**

I. INTRODUCTION

Super resolution is a technique to generate high-resolution image given a single or multiple low-resolution images. This process is needed and is becoming reconstruct in many digital image and video processing applications especially video surveillance. In most video surveillance application, identifying suspect′s face is very important. However, the majority of video surveilance cameras delivers low resolution image, low lightning condition and camera angle position also contribute to the quality of the images. Thus, super resolution methods as a post processing method for low resolution images are becoming needed these days.

There are three categories in image super-resolution: interpolation based methods, reconstruction based methods, and learning based methods. Interpolating is simple method but it usually has poor performance especially blurred edge and lacking image details. Extracting a single high-resolution image from more multiple low resolution image, or known as reconstruction method, is not so efficient when applied to very low resolution face image without consideration on special characteristics of face images. In this paper, we focus on generating high-resolution face image from a single low-resolution image using learning based approach. We use this method in order to enhance the detailed characteristic of facial information. The algorithm can learn fine details in both low and high-resolution images. Various methods have been proposed by many researchers. For example, Freeman *et. al.*[1] proposed example-based approach on enlarging single image. In their training set, they stored corresponding pairs of patches: the band-pass or high-pass filtered and contrast normalized. Other researchers [2], [3], [4] proposed neural network approach to generate high resolution image. Researchers [5] also proposed super-resolution approach based on its sparse representation.

Face hallucination is another popular approach in face super-resolution. Face hallucination was originally proposed by Baker and Kanade [6] but the gradient pyramid based prediction cannot model the face priors very well because the pixel are predicted individually causing artifacts and discontinuity. Later Jianchao Yang *et. al.*[7] proposed face hallucination via sparse coding.

Representing image using sparse coding is very attractive approach at present. Sparse coding proposed by researchers [8] is very efficient. It can learn basis functions that capture higher-level features in the data, even only given unlabeled input data. Researchers [9] used these approaches to train dictionary in order to generate high resolution from single low resolution image. To solve penalty of the high patch prediction, they used least absolute shringkage and selection operator (Lasso) adopted from Least Angle Regression [10].

In this paper, we propose method to generate high-resolution face image given single low-resolution image using sparse representation and get the solution using Elastic-Net, [11] an improvement of Naïve Elastic Net [12].

The remainder of this paper is organized as follows. In section II, we explain about association between sparse representation and super resolution. Ridge regression, Lasso, and Elastic Net, solutions to sparse coding, are described in section III. Section IV shows the result and analysis. This paper gives conclusion in the last section.

II. SPARSE REPRESENTATION FOR SUPER RESOLUTION

In simple words, sparse representation is a representation that describes most or all information from a signal using linear combination of a small number of element from the signal. These elements, or called atoms, are chosen from over-complete dictionary. Sparse coding is a way to find a representation with a small number of significant coefficients.

A. Sparse Representation

How to represent signal into sparse representation has been summarized by researchers [13]. Suppose D $\in \Re^{n\times k}$ is an over-complete dictionary can be represented in sparse representation, as in (1).

$$y = LD\beta_0 \tag{1}$$

Where vector $\beta \in \Re^k$ contains the representations coefficients of the signal y and $L \in \Re^{k\times n}$ is a projection matrix. The sparest representation is the solution of either

$$(P_0) \min_{\beta} ||\beta||_0 \quad s.t. \quad y = LD\beta \tag{2}$$

or

$$(P_0, \epsilon) \min_{\beta} ||\beta||_0 \quad s.t. \quad ||y - LD\beta||_2 \leq \epsilon \tag{3}$$

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. At first step, it finds one atom that best matches the signal, and finds the next one. Orthogonal Matching Pursuit is an improvement of Matching Pursuit.

The second way, Basis Pursuit, suggests a convexication of the problems mentioned in (2) and (3), by replacing the ℓ_0-norm with an ℓ_1-norm. This approach can be motivated based on Maximum A Posteriori (MAP) estimation. In this paper, we use Basis Pursuit to solve super-resolution problem, because of the advantages of such convex method are its stability and high estimation accuracy.

B. Super Resolution via Sparse Representation

Jianchao Yang *et. al.* [5] explained how to represent super resolution process via sparse representation. If we talk about super resolution, we consider how to generate high resolution image $\boldsymbol{X}$, given low resolution image $\boldsymbol{Y}$. Because of low-resolution image is down sampled from high-resolution image by factor $\boldsymbol{S}$ and blurred by $\boldsymbol{H}$ blurring filter. We can represent the reconstruction constraint, as in (4).

$$Y = SHX \tag{4}$$

Super resolution is still ill-posed, since much information of the high resolution block or patch is missing. To minimize the missing information, we divide an image into small patch y. Freeman *et. al.* [1] divided low resolution image into a small patch of 7×7 pixels. The overlap between adjacent high-resolution patches was 1 pixel. The patches were band filtered and contrast normalized. In spite of using band-pass filter, Jianchao *et. al.* [5] used first order and second order gradient of the patches as the representation.

Let $x \in \Re^n$ is a signal vector. From (1), signal x can be represented as $x = D\beta$. To infer the high resolution patch for its low-resolution one, researchers [5] create two dictionaries high-resolution patches, D_h and coresponding low-resolution patches, D_l, For each input low-resolution input patch, y, they found a representation with respect to D_l and its corresponding vector in D_h becoming the best matching high-resolution patch.

III. Solution to Sparse Coding

Sparse representation of the signal can be computed using regularized least square method. Convex penalties are considered to develop efficient algorithms for estimation of the generalized linear least square method. The penalities are ℓ_1 (the lasso), ℓ_2 (ridge regression), and combination of the two (elastic net).

A. Ridge Regression

One issue when we deal with Least Square regression is about prediction accuracy. It means that we have to minimize mean squared error by considering bias and variance. To solve the Ordinary Least Square, we have to get the smallest residual sum of squares.

Suppose we have training set $D = \{X_i, y_i\}_{i=1}^n$, the residual sum of square, $RSS(\beta)$, is shown in (5).

$$RSS(\beta) \equiv \sum_{i=1}^{n} (\hat{y_i} - X_i\beta)^2 \tag{5}$$

First improvement of the Ordinary Least Square is Ridge regression. The penalty of the ridge regression, $L_2(\beta)$ is shown in (6).

$$L_2(\beta) = \lambda \sum_{j=1}^{p} |\beta_j|^2, \lambda \geq 0$$
$$\beta_{ridge} = \underset{\beta}{\text{argmin}} ||\hat{y_i} - X_i\beta||^2 + \lambda_2 ||\beta||_2^2 \tag{6}$$

Where λ is a constant and larger λ means greater shrinkage.

B. Lasso

To find sparse solution, Lasso [14] is a method for regression that uses an ℓ_1 penalty, also known as Basis Pursuit[15]. Efron et. al. [10] proposed an efficient technique to solve the entire lasso coefficient at the cost of a full least-square fit.

Because the nature of ℓ_1 penalty, Lasso, as a variable selection method, continuously shrinks the coefficients to zero. Thus, the Lasso does a kind of continuous subset selection. To find the best correlated predictors, Lasso picks one and ignore the others so Lasso can produce a very sparse model. Lasso penalty can be calculated as shown in (7).

$$L_1(\beta) = \lambda \sum_{j=1}^{p} |\beta_j|^2, \lambda \geq 0$$
$$\beta_{lasso} = \underset{\beta}{\text{argmin}} ||\hat{y_i} - X_i\beta||^2 + \lambda_1 |\beta|_1 \tag{7}$$

Least Angle Regression, abbreviated LARS[10] is an approach that establishes the connection between the Lasso and Forward stage-wise regression. Suppose we have data consisting of *p* variables in *n* observations, the LARS algorithm continues to provide Lasso solutions along the way, and the

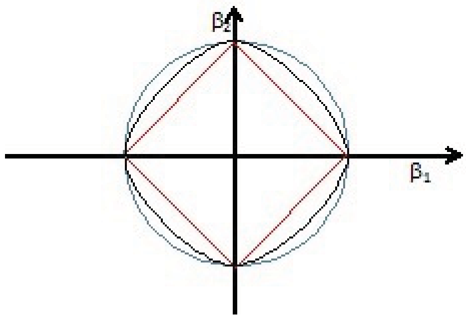

Fig. 1. Blue line is penalty of Ridge Regression, red line is penalty of Lasso, and black line is penalty of Elastic Net that compromise between two penalties.

final solution highlights the fact that a Lasso fit can have no more than mean centered variables with non-zero coefficients. If we have more variables than observation (p $\gg$ n), LARS terminates at the saturated least-squares fit after mean centered variables have entered the active set (at a cost of O(n^3) operations)[10].

C. Elastic Net

When we work with Image Processing, we find very larger variables than observation (p $\gg$ n). We will find tens of thousands of pixels included in less than two thousand blocks of 7×7 pixels. With this condition, we find disadvantage of Lasso. Because Lasso selects at most n variables before it saturates, and the most important thing to consider is Lasso cannot 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 [12].

Elastic Net is a new method that is powerful to solve this problem. Because Elastic Net includes Regression and variable selection, with the capacity of selecting groups of correlated variables. Naïve Elastic Net is compromised between Lasso and Ridge Regression. The optimization of Naïve Elastic net is shown in (8). Contour of the Elastic Net penalty compares with Lasso and Ridge Regression is shown in Fig.1.

$$\beta_{Na\ddot{\imath}veEN} = \underset{\beta}{\operatorname{argmin}} ||\hat{y}_i - X_i\beta||^2 + \lambda_1|\beta|_1 + \lambda_2||\beta||_2^2 \quad (8)$$

Naïve Elastic Net has combined ℓ_1 and ℓ_2 penalty where lambda 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. Naïve Elastic Net does not perform satisfactory because there are two shrinkage procedures (Ridge and Lasso). The procedures appear to incure the double amount of shrinkage. Double shrinkage does not help to reduce the variances much and introduces unnecessary extra bias, compared with pure lasso or ridge shrinkage. To achieve better performance, Naïve Elastic Net has to be rescaled and become Elastic Net[11].

The elastic net produces a sparse model with good prediction accuracy, while encouraging a grouping effect. The empirical results and simulations demonstrated by researcher [12] showed the good performance of the elastic net and its superiority over the lasso.

D. Elastic Net in Super-Resolution

Elastic Net as a solution in sparse representation of super-resolution is new approach since this method has not been proosed by other researchers. Super resolution via sparse representation was originally proposed by Jianchao *et. al.* [7]. The proposed approach used Lasso as a solution of the representation and gave better results than other conventional methods. Because of the number of variables is greater than the number of observations, Lasso does not give optimum solution of sparse representation. Lasso selects one variable among a group of highly correlated variables and does not care which one. Elastic net will remove the limitation by compromising with Ridge constraint. With the grouping effect on the elastic net, the reconstructed high resolution image will be expected to have a lower error and more accurate. This grouping effect is very useful to get information from the dictionary to generate high resolution image.

Algorithm

1. Input: sparse basis matrix X, training dictionaries D_h and D_l, a low resolution image Y
2. Upsize Y using Bicubic interpolation
3. For each patch 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 (8):
   $$\beta_{EN} = \underset{\beta}{\operatorname{argmin}} ||y - \beta X||^2 + (1-\alpha)\lambda_2||\beta||_2^2 + \alpha\lambda_1|\beta|_1$$
   b. $\beta*$ = Normalized Coefficients β
4. Output: super-resolution face $X*$

The algorithm explains how to apply Elastic Net to solve sparse representation of the super-resolution process. The main idea is to compromise between Lasso and Ridge Regression. We use factor α to compromise the value of λ_1 and λ_2 in (8). In this paper, this algorithm is tested using frontal face image view.

IV. Experimental Results

In this section, we will discuss about our experiment applying Elastic Net in super-resolution process. Experiment in this paper is based on Jianchao *et. al.* [7], and then to find solution from sparse algorithm, we used Elastic Net from Glmnet MatLab package whose algorithm made by J. Friedman *et. al.*[11] to improve Lasso which based on Least Angle Regression by Efron *et. al.*[10].

A. Database

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 45 of them having 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 × 7 and overlaps 1 pixel with adjacent patch. It gives size of D_h 49 × codebook size. In order to extract different features for the low resolution image patches, which size is 3 × 3 (upsampled to 6 × 6), be applied first-order and second-order derivatives. The four $1-$D filters is used to extract the derivatives are:

$$f_1 = [-1\ ,\ 0\ ,\ 1] \quad f_2 = f_1^T,$$
$$f_3 = [-1\ ,\ 0\ ,\ -2\ ,\ 0\ ,\ 1] \quad f_4 = f_3^T \tag{9}$$

Applying those filters give size of D_l 144 × 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].

B. Glmnet

Glmnet is a General public License implemented in the R programming system (R Development Core Team 2009) which is available in Tibshiranis web page [17]. Glmnet is an efficient algorithm for fitting the entire lasso or elastic-net regularization path for linear regression, logistic and multinomial regression models. The algorithm uses cyclical coordinate descent in a pathwise fashion, based on papers [11], [18]. This package consists of main glmnet code, prediction, print, plot, code to find coefficients, and cross validation. This package also includes Fortran code to find sparse solution. We use parts of glmnet MatLab package to evaluate Elastic Net in this research. By some modification in the parameters, we could get results of solving sparse coding by Elastic Net in this super-resolution process.

C. Results and Discussion

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 × 40 pixels and four times to 30 × 30 pixels. Even we test 45 images; in this paper, we only show the results of 5 different images, ordinary, eyeglasses, mustache, dark color, and smiling face. Fig. 2 shows that even in small size of dictionary, 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 1 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.

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 II, 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. 2.

To prove the better performance of Elastic Net, we try to zoom low resolution images four times. At first, a high-resolution image, which size is 120 × 120 pixels, is downsampled four times to 30 × 30 pixels. The low-resolution image is very small; we lose much information during downsample process. It is more difficult to upsize the resolution and

TABLE I
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 II
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

TABLE III
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

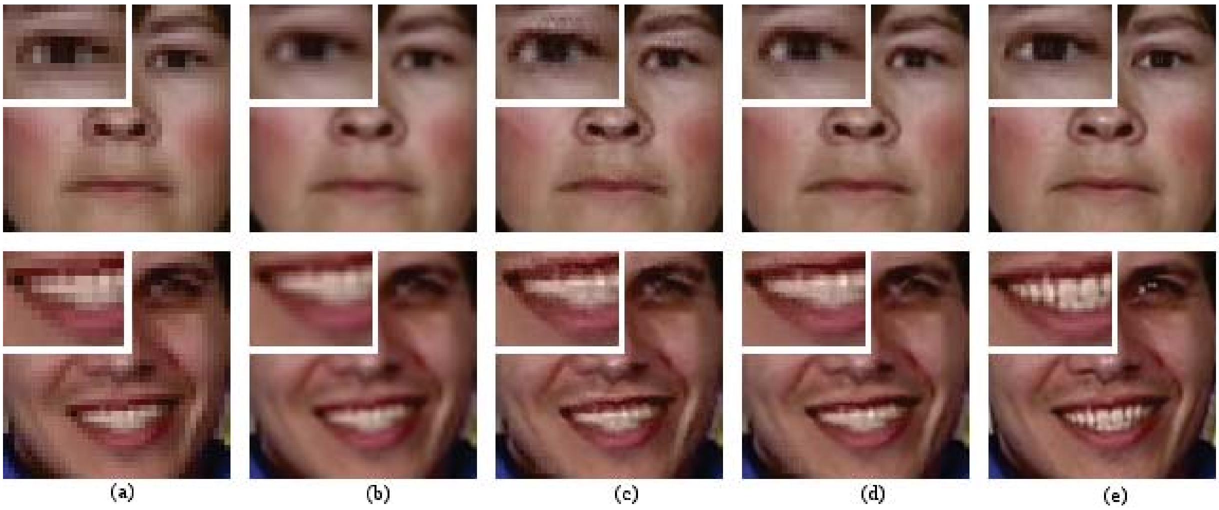

Fig. 2. High-resolution image generated by Lasso and Elastic Net in various zooming factor: (a) Lasso zooming factor 4, (b) Elastic Net zooming factor 4, (c) Lasso zooming factor 3, (d) Elastic Net zooming factor 3, (e) original image.

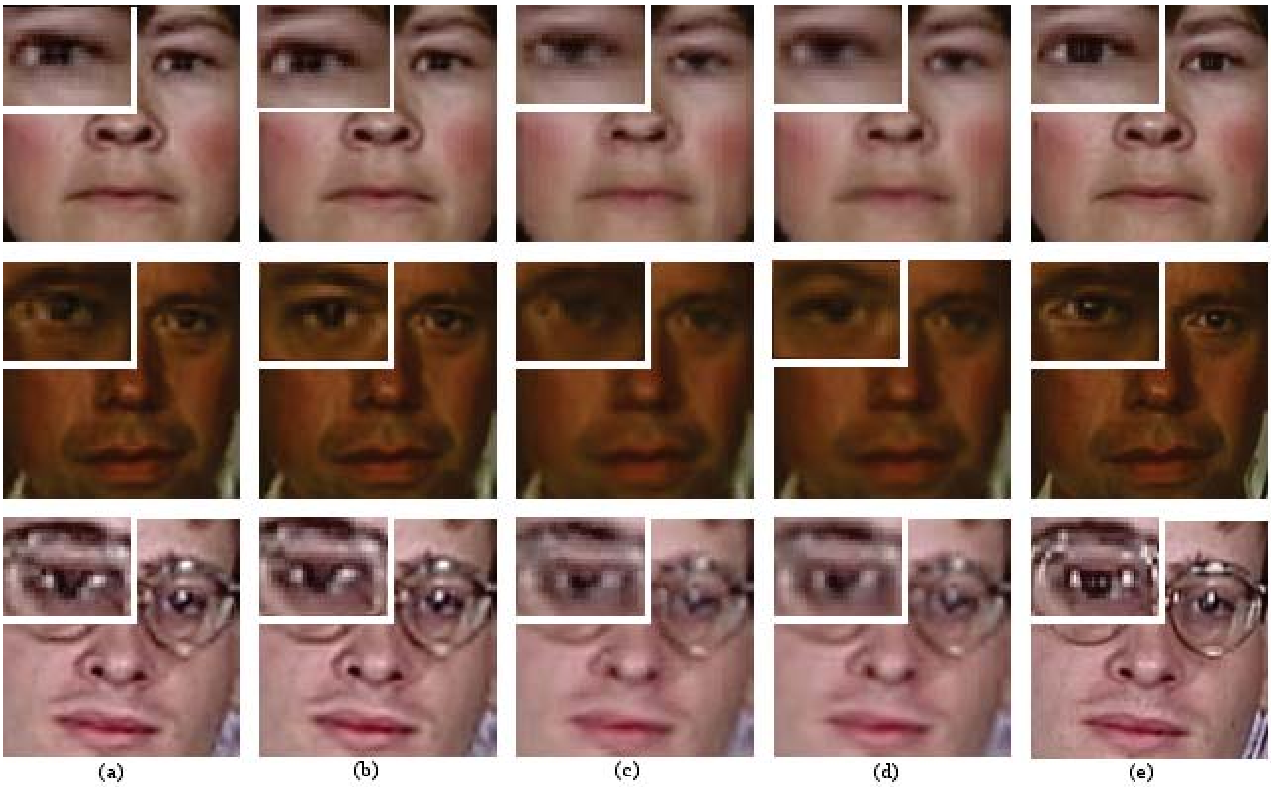

Fig. 3. High-resolution image generated by Lasso and Elastic Net in various zooming factor: (a) Lasso zooming factor 3, (b) Elastic Net zooming factor 3, (c) Lasso zooming factor 4, (d) Elastic Net zooming factor 4, (e) original image.

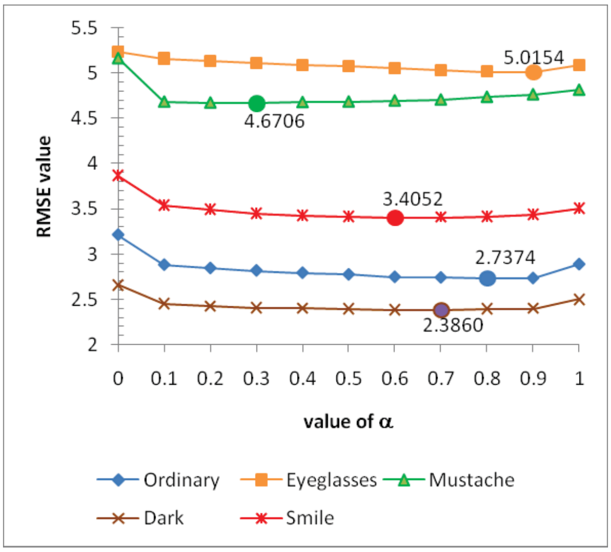

Fig. 4. RMSE of generated high-resolution image by various value of α.

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 is presented in Table III and Fig. 3.

As we see in Table IV, 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. 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.

V. Conclusions

Super resolution is a kind of process which a number of variables are very larger than a number of observations.

TABLE IV
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

Elastic Net can do some group selection when generate coefficients and compromising between Lasso and Ridge Regression. 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.

Acknowledgment

This research has been supported the collaborative research project titled Video Processing and Transmission, in JICA Project for AUN/SEED-Net, Japan.

References

[1] W. T. Freeman, T. R. Jones, E. C. Paztor, "Example Based Super-Resolution," *IEEE Computer Graphics and Applications*, vol. 22, 2002.
[2] C Marivet and F. B. Rodriguez, "A two-step neural-network based algorithm for fast image super-resolution," *Image and Vision Computing*, vol. 25, pp. 14491473, 2007.
[3] V. H. Patil, D. S. Bormane, V. S. Pawar, "Super Resolution using Neural Network," *Second Asia International Conference on Modelling and Simulation*, 2008.
[4] Y. Huang and Y. Long, "Super-resolution using neural networks based on the optimal recovery theory," *Journal of Computer Electronics*, vol. 5, pp. 275281, 2007.
[5] J. Yang, J. Wright, T. Huang, Y. Ma, "Image Super-Resolution as Sparse Representation of Raw Image Patches," *Proc. CVPR*, 2008.
[6] S. Baker and T. Kanade, "Hallucinating faces.," *IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.
[7] J. Yang, H. Tang, Y. Ma, and T. Huang, "Face Hallucination via Sparse Coding," *IEEE International Conference on Image Processing (ICIP)*, 2009.
[8] H. Llee, A. Battle, R. Raina and Andrew Y. Ng, "Efficient Sparse Coding Algorithm," *Proceeding of Neural Information Processing System (NIPS)*, 2007.
[9] J. Yang, J. Wright, T. Huang, Yi Ma, "Image Super-Resolution via Sparse Representation," *submitted to IEEE Transactions on Image Processing*, 2009.
[10] B. Efron, T. Hestie, R. Tibshirani, "Least Angle Regression," *The Annals of Statistics*, vol. 32, pp. 407−499, 2004.
[11] J. Friedman, T. Hestie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, vol. 33, 2010.
[12] H. Zou and T. Hestie, "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society B*, vol. 67, pp. 301−320, 2005.
[13] M. Aharon, M. Elad, A. Bruckstein, Y. Katz, "K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation," *IEEE Transaction on Signal Processing*, vol. 54, 2006.
[14] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, pp. 267-288 1996.
[15] D. Donoho, Chen SS, and M. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, vol. 20, p. 33, 1998.
[16] A. V. Nevian, "Georgia Tech face database," 1999.
[17] R. Tibshirani, "Glmnet for Matlab " 2010.
[18] J. Friedman, T. Hestie, H. Hofling, and R. Tibshirani, "Pathwise coordinate optimization," The Annals of Applied Statistics, vol. 1, pp. 302−332, 2007.