Face super-resolution algorithm based on SVM-improved learning

Jiali Tang[1, 2*], Chenrong Huang[3], Yijun Liu[2], Honghui Fan[2], Jianmin Zuo[1, 3]

[1] *School of Mechanical Engineering, Jiangsu University, Zhenjiang 212013, Jiangsu, China*

[2] *College of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, Jiangsu, China*

[3] *School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, Jiangsu, China*

Received 1 June 2014, www.cmnt.lv

Abstract

As many other inverse problems, human face image super-resolution is an ill-posed problem. The problem has been approached in the context of example-based superresolution learning. However, these methods need to run though all the sample set, which results in high calculation load and image degradation because of mis-matching. In this paper, we propose a new face image superresolution algorithm based on Support Vector Regression (SVR) pre-classified learning. A Principal Component Analysis (PCA) based pre-process is used to select a subset of samples. Then the best-matching sample images are trained to ensure the content relevance between the sample patch and the input low-resolution image. Further improvement involves a combination of classification and SVR-based techniques. Therefore, experiment results show that the proposed algorithm gets better reconstruction performance and faster program running speed.

Keywords: Face Super-resolution, Support Vector Machine (SVM), Principal Component Analysis (PCA), Example-based Algorithm

1 Introduction

The super-resolution reconstruction is the technology of using software algorithms to obtain higher resolution images without changing current imaging system hardware, which has been an active research in image processing and is widely applied in medical diagnosis, remote sensing and many other areas. Among them, face image has been specially focused in the field of image processing and pattern recognition. The traditional face super-resolution algorithms usually accord to the multiframe low-resolution (LR) images of sub-pixel displacement under same scenes to reconstruct high-resolution (HR) images. The single frame face image super-resolution algorithm based on training set reconstructs the high-frequency details of images by learning the relation between HR and LR images, which has become the mainstream of research area in recent years [1-5].

Human face super-resolution algorithm was first proposed by the Carnegie Mellon University's Simon Baker [6] etc., which builds a Gauss Pyramid with multi-resolution feature of HR images. By matching in different resolution spaces in the Pyramid and using the feature of images, the algorithm searches the corresponding HR patch from the side of matching the input image's feature. Framework of Bayesian inference is used to build the generating optimized model of HR images, then reconstructs the HR face images. This algorithm gets better reconstruction result than interpolation and traditional methods based on the Markov Network model. However, it is necessary to set up a huge sample database where we can do the traversal search, which leads to a heavy workload of calculating and the mismatching in search result will cause final quality decrease badly. Furthermore, the algorithm doing super-resolution reconstruction relies on its Gauss Pyramid model from the model of matched feature. It is limited for the improved room of subjective reconstructed quality when there is aliasing in sample feature extracting.

In this paper, we propose a face super-resolution algorithm based on SVM-improved learning. To solve the problems based on learning-based algorithm such as the heavy work of calculating and mismatching, we use a classifier to pre-classify the sample database and select the subset that is similar to the object's PCA feature, which reduces the mismatching rate of sample patch, shortens the program's running time and improves the image reconstruct performance.

2 SVM learning

SVM is on the base of Vapnik-Chervonenkis (VC) Dimension theory and Structural Risk Minimization (SRM) principle, to get better generalization ability according to limited sample details to find a best way to balance between the complexity of model and learning ability [7]. The main idea of SVM is to build a linear hyperplane as the decision boundary to make the margin

* ***Corresponding author*** e-mail: tangjl@jsut.edu.cn

between positive example and negative example maximum. Please see Figure 1.

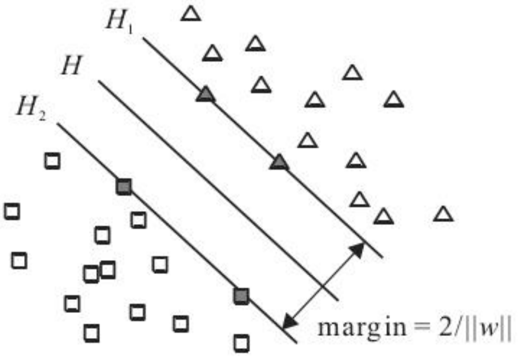

FIGURE 1 Schematic diagram of SVM learning.

More exactly to speak, SVM is a kind of generalized linear classifier based on SRM principle whose basic strategy is to confirm the empirical risk and to make confidence risk minimum.

The general type of classification hyperplane can be written as $wx+b=0$, after normalized, which makes the linear separable sample set $(x_i, y_i), x_i \in R^d, y_i \in \{+1,-1\}, i=1,2,\cdots,n,$ meeting the following requirement:

$$y_i(wx_i+b) \ge 1, i=1,2,\cdots,n \,. \tag{1}$$

When margin is equal to $2/\|w\|$, so it is equal between maximizes margin and minimizes $\|w\|^2$. A hyperplane, which fits the condition (1) and makes $\frac{1}{2}\|w\|^2$ minimum is named optimal classification hyperplane, the training sample set is called support vectors, which makes the equal sign exist. Bring in the Lagrange function:

$$L = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} a_i y_i (w \cdot x_i + b) + \sum_{i=1}^{l} a_i \,. \tag{2}$$

Among in Equation (2), $a_i > 0$ is the coefficient of Lagrange. Transform the question about optimal classification hyperplane into its dual problem by Lagrange optimization method, namely means becoming the question under the constraint of the inequality to search the best answering whose answer is unique solution. The optimal classification function is:

$$f(x) = \mathrm{sgn}\left\{\left(w^* x\right) + b^*\right\} = \mathrm{sgn}\left\{\sum_{i=1}^{n} \alpha_i^* y_i (x_i x) + b^*\right\}. \tag{3}$$

Among Equation (3), α_i^* is the corresponding Lagrange multiplier of each sample, we can prove only a few α_i^* is non-zero, then the corresponding sample is support vector. b^* is the classification threshold, which can acquire by any support vector or averaging from any couple of support vector.

For the sample set linearly non-separable case, we can put a slack variable $\delta_i \ge 0$, it becomes:

$$y_i(wx_i+b) \ge 1-\delta_i, i=1,2,\cdots,n \,. \tag{4}$$

And change the target into:

$$\min_{w,b,\delta} \varphi(w,x) = \frac{1}{2} w^T w + C\sum_{i=1}^{n} \delta_i \,. \tag{5}$$

To get the generalized optimal classification hyperplane by building a soft margin that considering in half way between minimum mistakes sub-sample and maximum margin, $C > 0$ is a constant, which controls the level of punishment for the wrong sub-sample.

For the non-linear case, we can change it into a high-dimensional linear problem by non-linear variation. Generally speaking, this kind of non-linear variation is more complex, and hardly come true through the algorithm. We can use kernel function $K(x_i, x_j)$ who meets the Mercer's condition instead of original space's inner product, which will realize the linear classification after some non-linear variation, so that will avoid the concrete form of non-linear variation. The classification function is becoming:

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x_j) + b^*\right). \tag{6}$$

The generator classification function is used to pre-classify the data in images training base, so as to judge the correlation between the training images and target images.

3 Face super-resolution reconstruction based on SVM-improved learning

Face super-resolution reconstruction means to estimate HR face by LR face, which is an application of imaging super-resolution technology in face area. As a kind of special image, due to the conspicuous structural features in face images, so the super-resolution algorithm adapt to common images can get a better effect in face reconstruction by combining these priori features. It is also widely applied in computer visual video surveillance, such as the field of video surveillance, usually the location of camera is quite far away from the face, so the captured face images are quite small. In order to better identify images, we first enlarge the face by super-resolution, then identify them, or doing the enlargement and identification at the same time.

Existing example-based learning face super-resolution reconstruction method uses the relation between the LR images and corresponding HR images in Markov network learning sample-base, then restoring the high-frequency details of input LR images through the relation we learned. This algorithm needs to build a huge face sample-base and do the traversal search in it. The HR sample can be the high-frequency details of the LR sample to restore the face. If and only if the searched face sample is similar to the input image sample and there is continuity in the content between corresponding HR sample and other HR samples whose spatial position is contiguous, this HR sample can be the high-frequency details of the LR sample to restore the face. The searching process of the large-scale training sample base not only causes the time-consuming but also the mismatching of the search result will reduce the final quality.

According to the problems above, this paper proposes a face super-resolution algorithm based on SVM-improved learning. According to the PCA feature of human face, images are pre-classified by the algorithm. Before the matching search, SVM will pre-classify the sample base. So as to effectively extract and reconstruct the sub sample base, which has similar feature in target PCA, the algorithm establishes a SVM predictor for each class sample. By training samples, the algorithm trains the predictor and restores the prior knowledge in the form of predictiors' parameters. The experiment proves that it is an effective face super-resolution restoring method.

3.1 FACE FEATURE EXTRACTION

Face recognition technology has a wide application prospect in computer vision, guest identity authentication, multimedia data retrieval and etc.; face feature extraction is a necessary key step in process of face recognition. The face feature, which is related to application area, can be quickly and effectively extracted; becoming a key point whether realizes face automatic identification. The extraction process of human face features is a basis on a process that a transformation matrix maps human face image vector into feature vectors. During the mapping process, identification information is closely related with the application domain (such as different types of facial expression) can be extracted. And a large number of other irrelevant information (such as eyebrows shade, mouth size etc.) are discarded. In order to turn human face vector into human face feature vector with transformation matrix, it uses large quantities of training sample (known as the expression of face image type) to extract algorithm from human face feature. Then more training samples better the feature extraction.

PCA (Principal Component Analysis) is also called K-L transform (Karhunen-Loeve transformation), this method is to find a subset of main components which has the statistical distribution's data set at random, and corresponding base vectors satisfy the orthogonality [8]. The original data set is transformed into principal component space, which makes the cross-correlation of single data samples decrease to the lowest point. PCA is a classical method of feature extraction and data representation; it is widely applied in pattern recognition and computer vision, and becomes one of the most successful face recognition methods. Due to the dimension of face image sample is very high, if we directly process face images, the amount of calculation is large and the running time is long, so in recent years many face detection and face recognition algorithms use PCA method to reduce dimension first, then use other methods to process more.

A face image with $m \times n$ is rearranged as a column vector with $m * n$ dimension by PCA method, then all of the training images can get a group of column vectors after that change: $\{xi\}, x_i \in R^{m+n}, i = 1, 2, \ldots, N$, among that, N represented the number of the images in training sample set. Treat images as random column vectors, and training sample set is utilized to evaluate the average vector μ and covariance matrix S_T :

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \ , \tag{7}$$

$$S_T = \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T = XX^T, X = (x_1, x_2, \cdots, x_n) \, . \tag{8}$$

Choose the projection matrix A corresponding to the first k and biggest eigenvalues of S_T . Equation (9) is utilized to reduce the dimension of original images:

$$y = (X - \mu) A_K \, . \tag{9}$$

After the eigenvectors of S_T are restored as image matrix, which is the standardized face. The basic idea of PCA is that approximating each face by the linear superposition of standardized face. And let these linear coefficients as face features, these features are used to classify. The way to classify face with the use of PCA technology is called Eigenfaces method.

3.2 PRE-CLASSIFICAITON OF SVM

If there are K kinds of images in the images training database, mark as $T = \{P1, P2, P3 \ldots Pk\}$, and have k semantic classifiers $\{C1, C2, C3 \cdots Ck\}$. For each SVM classifier, its training set $T = \{(x_1, y_1), (x_2, y_2) \cdots (x_n, y_n)\}$, (x_1, y_1) is the labelled sample image which is given beforehand. $x_i \in R^2$ is the PCA feature of images; $y_i \in (-1, +1)$, +1 means images contain PCA semanteme, −1 means don't. Achieve images' semantic classifiers by using SVM to train these samples, and then these semantic classifiers are utilized to distinguish the left un-labeled images.

In this paper, using LibSVM 3.0 software package developed by Professor Zhiren Lin [9] of Taiwan University as SVM classify platform. LibSVM mainly includes *svmscale*, *svmtrain* and *svmpredict* three functions. Among that, *svmtrain* turn to train samples, according to input vectors and specified classification to generate suitable classifiers; when the way of classification is unknown, *svmpredict* is used to make input vectors map to the inner vector space of models based on specified classifiers, to find its classification principles; *Svmscale* is mainly for regularization of numerical value, numerical value regularize input vectors to the boundary where LibSVM is fit for processing.

Due to the diversity of kernel function, one of the main points of designing SVM is that choose suited kernel function and corresponding parameters. According to the research by Vapnik and others, different kernel functions have little influence on the performance of SVM, but it is crucial for the performance of support vectors classifiers to choose penalty coefficient and corresponding paremeters of kernel function. In this paper, we use radial basis function, which is widely applied:

$$K\left(x, x_i\right) = \exp\left(-\gamma\left|x - x_i\right|^2\right). \tag{10}$$

Net check search is used to cross-validation C and γ. Try each basic couple (C, γ), and then choose the couple of parameters with the highest accuracy rate during the cross-validation. During the cross-validationn over and over again, determine two parameters, based on 1-a-1 multi-classification method to train image samples.

3.3 SVM PARAMETER OPTIMIZATION AND TESTING HR IMAGES GENERATION

In RBF-SVM, we need to determine two parameters which are planning factor C and Gaussian width σ. For C, if $C \to \infty$, means the rules of classification will meet all of constraint conditions, not only enhance training time and complexity of training, but also make classification have over-learning's situation, which reduce its generalization performance. So the data range of C should meet the situation of TAR and TRR, in order to keep the generalization performance of classifier, we need to take value as small as possible. For the Gaussian width σ, if without some ask for algorithm, we just set a group of probable values through the experience, and solving in turn under σ, then adjust according to the answer.

Do the SVM parameter optimization through net check search method. Net check search is reached through specially quantize and partly optimize method of exhaustion. In the method of net check search, firstly, confirm the data range of parameters, and then confirm the step of take value. Each parameters can start from the beginning of take value, take value to the end according to step, after enumerating all of parameter combination, then based on corresponding evaluation criterion to assess classification result, to find the optimum parameter.

When the classification is completed, the algorithm will establish groups of predictor corresponding with samples in each region, each category corresponding predictor to that class, i.e. if the samples divided into several K class, the groups of predictors in this region ought to contain the predictors of this K class. Multiclass predictor consists of a group of sub predictor; each sample of base classes corresponds to a linear sub predictor. A group of predictors operate the combined prediction to construct multiclass predictors. When predicting, firstly input data coding by the generator codebook in the process of sample classification, namely classification. Then according to the category, the system selects the corresponding sub predictors to predict. The training process of multiclass predictors is the process of training each single sub predictor. After the completion of training the predictors, the parameters will be stored for guiding the super-resolution restoration which inputting low resolution image. After the completion of training multiclass predictors, the system will use the multiclass predictors to reconstruct the HR face image. By using the bilinear interpolation to magnify HR face image size, the system inputs the LR face image, which is used as the estimation of initial image, for super-resolution restoration. When inputting low resolution block, according to the region which it belongs, the corresponding predictors operate high frequency prediction. The high frequency information blocks, which produced by those predictors, will be used to add up to the estimation of initial images in order to form the outcome of HR face image, namely the realization of face image super resolution restoration.

4 Simulation result and analysis

Text images come from a subset of international standard face image database, including 200 people's images from 7 angers and 1400 pictures in all, which includes location parameter of eyes, noses, mouths and etc. for each face image. Before the experiment, do the image intensity normalization and geometric normalization (80×80 pixels per image).

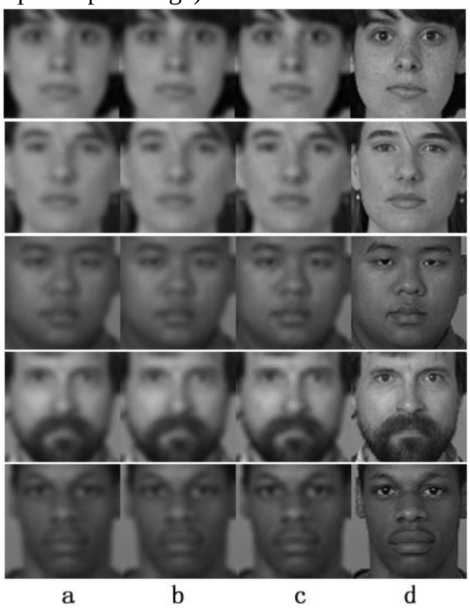

FIGURE 2 Reconstruction effects of three algorithms.
(**a** Freeman, **b** Reference [5], **c** Our method, **d** Original HD image)

Randomly choose 190 front view face images as training sample, PCA eigenvector is used as the input vector of support SVM, learning for the image categories, pre-classify again based on output result. We choose the left 10 front face images in the database as text images, use a Gaussian blur template with the size 5×5 and variance $\sigma = 0.85$ to degrade all HR images to the LR images with 40×40 pixels through Gaussian blur and down-sampling.

Do the contrast experiment apart with Freemam algorithm and document algorithm. We test the five face images' reconstructed effect (Figure 2).

As shown in these images, Freeman's algorithm [10] is the simplest, but the effect is poor. The algorithm of reference [5] is better in the central area of human face, but the other part especially the edge of face is fuzzy. Our approach is better than the first two algorithms in visual effect, but the edge of face and hair line still have some masks.

The value of PSNR is used to compare then text images' objective effect. Generally speaking, the higher value of PSNR is, the less distortion of image is. See Equation (11).

$$PSNR = 10 \times \lg\left[\frac{\left(2^n - 1\right)^2}{MSE}\right]. \tag{11}$$

In Equation [11], *MSE* represents the error of mean square between original images and processing images.
The result of quantizing PSNR is shown in Figure 3. We can reach the conclusion that the PSNR values in our approach basically higher than the other two algorithms. What's more, calculating apart the running time of image super-resolution reconstructed procedure on Freeman and our algorithm by *tic* and *toc* function in Matlab, the running speed of this paper enhanced obviously.

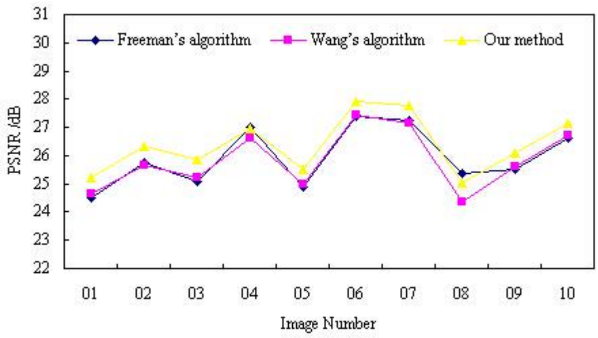

FIGURE 3 PSNR comparison of three algorithms

5 Conclusions

In this paper, SVM algorithm has been applied to example-based human face super-resolution reconstruction, which is a new attempt. SVM is used as the training classifier because its sample learning ability is very strong. We pre-search the sample database by defining PCA feature, find the sub-sample database whose feature is similar, then do the pixel-level exact matching search in the sub-sample set. The experiment shows that our approach can effectively reduce the mis-matching of image patches, save the running time and reach a higher quality in the reconstruction. In both subjective and objective way, restored image is better than the classical insert algorithm and the learning-based super-resolution method, and in the meantime the program running time is effectively improved.

Acknowledgments

The authors are indebted to their colleagues of School of Computer Engineering, Jiangsu University of Technology, for their suggestive discussion and zealous help. This study is supported by the National Natural Science Foundation of China (61302124), and the IUR Prospective Joint Project of Jiangsu Province (BY2014038-05), and also the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (13KJB520006).

References

[1] Gunturk B K, Batur A U 2003 Eigenface-domain Super-resolution for Face Recognition *IEEE Transactions on Image Processing* **12**(5) 597-606
[2] Huang H, He H T, Fan X 2010 Super-resolution of Human Face Image Using Canonical Correlation Analysis *Pattern Recognition* **43**(7) 2532-43
[3] Hu Y, Lam K M, Shen T Z 2011 A Novel Kernel-based Framework for Facial-image Hallucination *Image and Vision Computing* (29) 219-29
[4] Jia K, Gong S G 2008 Generalized Face Super-Resolution *IEEE Transactions on Image Processing* **17**(6) 873-86
[5] Wang X G, Tang X O 2005 Hallucinating Face by Eigentrans-formation *IEEE Transactions on Systems Man and Cybernetic* **35**(3) 425-34
[6] Baker S, Kanade T 2000 Hallucinating Faces *Proc. Conf. on Automatic Face and Gesture Recogniton* Grenoble, France 83-8
[7] Valentini G, Dietterlch T G 2004 Bias-variance Analysis of Support Vector Machines for the Development of SVM-Based Ensemble Methods *Journal of Machine Learning Research* (5) 725-75
[8] Yang J, Zhang D 2004 Two-dimensional PCA: a new approach to appearance-based face representation and recogniton *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(1) 131-8
[9] Chang C C, Lin C J 2011 LIBSVM: A library for support vector machines *ACM Transactions on Intelligent Systems and Technology* **2**(3) 1-27
[10] Freeman W T, Jones T R**, Pasztor** E C 2002 Example-based super-resolution *IEEE Computer Graphics and Applications* **22(2) 56-65**

Authors

Jiali Tang, born on May 10, 1980, China

Current position, grades: associate professor at Jiangsu University of Technology, China.
University studies: Msc. Degree in Material Processing Engineering from Nanjing University of Aeronautics and Astronautics, China in 2005.
Scientific interests: As an IEEE member, his research interests include industrial control software development and image super-resolution reconstruction. He has published nearly twenty academic papers in related journals.

Yijun Liu, born on June 28, 1978, China

Current position, grades: works in School of Computer Engineering of the Jiangsu University of Technology.
University studies: bachelor's degree in Computer Science and Technology from Nanjing University in 2000, her Master's degree in Computer Science and Technology from Nanjing University in 2003.
Scientific interests: machine learning, data mining and intelligent information system.

Honghui Fan, born on October 24, 1980, China

Current position, grades: is a member of China Computer Federation. In 2011 joined Jiangsu University of Technology.
University studies: M.Sc. and Ph.D. from the Yamgata University of Japan in 2008 and 2011, respectively.
Scientific interests: ultrasound imaging, image restoration, digital image, and signal processing in biomedical engineering.

Jiamin Zuo, born on January 7, 1957, China

University studies: PhD supervisor in the School of Mechanical Engineering of Jiangsu University. **Current position, grades:** professor at Nanjing Institute of Technology, China.
Scientific interests: As a senior member of Chinese Mechanical Engineering Society (CMES), his research interests include electromechanical control and industrial automation.
Publications: over seventy academic papers in related journals.