IEEE Int. Conference Neural Networks & Signal Processing
Zhenjiang, China, June 8~10, 2008

AN EFFICIENT EXAMPLE-BASED APPROACH FOR IMAGE SUPER-RESOLUTION

Xiaoguang Li[1,2], Kin Man Lam[2], Guoping Qiu[3], Lansun Shen[1] and Suyu Wang[1]
1. Signal & Information Processing Lab. Beijing University of Technology, Beijing, China, 100124
2. Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong
3. Department of Computer Science, Nottingham University, UK
lxg@emails.bjut.edu.cn, enkmlam@polyu.edu.hk

ABSTRACT

A novel algorithm for image super-resolution with class-specific predictors is proposed in this paper. In our algorithm, the training example images are classified into several classes, and each patch of a low-resolution image is classified into one of these classes. Each class has its high-frequency information inferred using a class-specific predictor, which is trained via the training samples from the same class. In this paper, two different types of training sets are employed to investigate the impact of the training database to be used. Experimental results have shown the superior performance of our method.

Key Words —— Example-based Super-resolution, Human face magnification, Class-specific predictor

1. INTRODUCTION

Image super-resolution plays an important role in many multimedia applications. This term refers to the reconstruction of a high-resolution (HR) image from a single or a set of low-resolution (LR) images [1]. In this paper, we consider image super-resolution based on a single image. This is also called image magnification or image interpolation. A number of super-resolution algorithms [2-5] have employed regularization terms to solve the ill-posed image up-sampling problem. However, using smoothness priors that are defined artificially has been found to lead to overly smoothed results [6,7]. Example-based or learning-based super-resolution algorithms [6-16] have been proposed recently as a very attractive approach for image super-resolution. Instead of defining a prior intuitively, this approach exploits the prior knowledge between the high-resolution and the corresponding low-resolution examples by learning algorithms.

Most example-based super-resolution algorithms [8-12] involve a training set, which is usually composed of a large number of HR patches and their corresponding LR patches. The input LR image is split into either overlapping or non-overlapping patches. Then, for each LR patch from the input image, either one best-matched patch or a set of the best-matched LR patches is selected from the training set. The corresponding HR patches are used to reconstruct the output HR image. Freeman et al. [8, 9] embedded two matching conditions into a Markov network. One is that the LR patch from the training set should be similar to the input observed patch, while the other condition is that the contents of the corresponding HR patch should be consistent with its neighbors. Wang et al. [10] extended the Markov network to handle the estimation of PSF parameters. Stephenson and Chen [11] presented a method in which the symmetry of a cropped human face is considered in the Markov network. Qiu [12] proposed an alternative method, based on vector quantization, to organize example patches. A survey of example-based super-resolution methods is available in [13].

The above-mentioned work has made significant contributions to the way we now exploit learning-based image super-resolution. However, most of these existing algorithms are only a kind of "searching and pasting" approach, and are therefore computationally intensive when searching for a LR-HR patch from a huge training set. Furthermore, best-matched but incorrect patches will seriously degrade the reconstruction results.

In this paper, we propose a new example-based super-resolution algorithm with a class-specific predictor so as to solve the above-mentioned problems in the existing algorithms. The main contributions of this paper are: (1) a class-specific predictor is designed for each class in our example-based super-resolution algorithm – this can improve the performance in terms of visual quality and computational cost; and (2)

978-1-4244-2311-8/08/$25.00 ©2008 IEEE.

different types of training set are investigated so that a more effective training set can be obtained.

2. OUR PROPOSED ALGORITHM

Although a scene from the real world contains an abundance of varied content, a small local block in an image can be classified into just a few categories, such as flat, edge, corner, and so on. In our algorithm, the classification is performed based on vector quantization (VQ), and then a simple and accurate predictor for each category, i.e. a class-specific predictor, can be trained easily using the example patch-pairs of that particular category. These class-specific predictors are used to estimate, and then to reconstruct, the high-frequency components of a HR image. Hence, having classified a LR patch into one of the categories, the high-frequency content can be predicted without searching a large set of LR-HR patch-pairs. The details of our algorithm are described in the following.

2.1. Generation of Training Databases

The training set selected for use is important to the performance of the example-based super-resolution methods. Each record in the training set is an example patch-pair, viz. a HR image block and the corresponding LR block. Similar to the method proposed by Qiu [12], a multi-resolution representation of an input image is formed using a three-level Laplacian Pyramid. Let I_0 represent a HR example image, which is blurred and down-sampled to produce I_1 by a zooming factor of z. Similarly, I_2 is generated from I_1 using the same zooming factor z. The up-sampled images from I_1 and I_2 are generated using bilinear interpolation with a factor z, and then subtracted from I_0 and I_1, respectively, to compute the difference images L_0 and L_1. The example patch-pairs are extracted from L_0 and L_1, which will then be used to train up the corresponding class-specific predictor.

For each block in L_0, there is a corresponding small block in the LR difference image L_1. If $z = 2$, each 4×4 HR block in L_0 has a corresponding 2×2 LR block in L_1. In order to maintain the continuity of a HR block with its neighbors, we extend the boundary of its corresponding LR block by 1 pixel to form a LR sampling block. This HR block and the corresponding LR sampling block thus form a patch-pair. By considering all the possible HR blocks in L_0 and the corresponding LR sampling blocks in L_1, a training set of patch-pairs is generated.

In this paper, we will consider two types of training set for example-based image super-resolution:

(A) Self-example training set (Set A): An input LR image is taken as the training image in the extraction of examples. The contents obtained from self-examples should be more relevant to the input image itself, and so the number of required training examples should be much smaller than that based on other images.

(B) Domain-specific training set (Set B): Images from a specific domain can be used to construct the training set. In this paper, we particularly consider facial images. Hence, the super-resolution of facial images based on our proposed algorithm will be evaluated. The training can be done off-line.

2.2. Content-Based Encoding/Classification

To infer the high-frequency information of an estimated HR image effectively, the original LR image is divided into patches, which are classified into different categories. Those patches belonging to the same category have similar texture characteristics. A predictor can be designed for each category in order to estimate the high-frequency content of the patches.

In our algorithm, VQ is used to encode an input patch. The number of levels or codevectors in the codebook is the number of categories to be used. In other words, each category is represented by a codevector. Hence, a codebook must first be trained based on either the input image for self-example training or a number of training images. Each training image I_0 is converted into images I_1 and I_2 by means of Laplacian decomposition, as follows:

$$I_1 = s_z(g(I_0)) \text{ and } I_2 = s_z(g(I_1)), \quad (1)$$

where $g(\ )$ and $s_z(\ )$ represent the Guassian operation and the sub-sampling operation with a factor of z, respectively. Then, the difference image L_1, which is the difference between I_1 and I_2, is constructed. This difference image is divided into a number of overlapping or non-overlapping blocks, and the corresponding HR blocks are then predicted. Following the work in [14], the block size is set at 4×4 in our implementation. The 16 elements of a block in L_1 are denoted as a vector, $b=[b_0\ b_1\ \cdots\ b_{15}]^T$, which is transformed to have zero mean and unit variance, as follows:

$$x = [x_0\ x_1\ \cdots\ x_{15}]^T, \quad (2)$$

where $x_i=(b_i-\mu)/\sigma^2$. μ and σ^2 are the mean and the variance, respectively, of the 16 elements b_i. With this normalization, the encoding or the classification of the blocks will become more efficient. Assume that all of the training vectors are classified into N different categories. Then, a codebook containing N codevectors has to be constructed. The codebook is denoted as

$$CB = \{\boldsymbol{c}_i \mid \boldsymbol{c}_i \in R^{16}, i = 0, 1, \ldots, N-1\}. \quad (3)$$

Vector quantization is employed for implementing content-based encoding, whereby the LBG algorithm [17] can be used for constructing the codebook. This codebook can be determined in advance or off-line, except in the case of training based on self-examples. In the encoding process, the best-matched codevector c_j to an input LR block is determined, and the index j represents the category of the input block. The corresponding j-th class-specific predictor will then be used to infer the high-frequency information.

All the training examples are encoded using the codebook. With the codebook for content-based encoding, each example patch-pair can be classified into one of the N categories. In other words, given a LR block of an example patch-pair after demeaning and normalization by (2), the closest vector is searched in the codebook. Then, the corresponding codevector is assigned to this patch-pair, where each codevector represents a category. Consequently, the training set is well structured with example pairs.

2.3. The Class-Specific Predictors

As described in Section 2.1, different training sets are generated, which are in the form of HR-LR patch pairs. Based on the LR part of the patch pairs, a codebook is trained so that each patch from a LR image can be encoded, and hence identified to belong to one of the N categories. In other words, with a given training set, the LR part of each training patch is classified by content-based encoding. Hence, each category contains a number of HR-LR training patches. Now, the remaining question is how to learn from these training patches to help the reconstruction of high-frequency information? In our algorithm, a class-specific predictor will be trained for each category. Upon training up the predictor for a category, the prior knowledge of HR-LR relations is stored in the weights of the predictor. This scheme achieves the goal of "learning" from the training examples, rather than just performing "search and pasting".

Figure 1 shows the implementation of our algorithm, which is composed of a content-based encoder to classify the input LR patches, and a set of N class-specific predictors. The well-known least-mean-squares (LMS) algorithm is used [18] to train up the predictors. The input to a predictor is the 4×4 blocks of the difference images L_1, while the output is the corresponding predicted HR blocks of the central 2×2 patches of the input blocks, as described in Section 2.1.

Note that the N class-specific predictors can be trained simultaneously. In the case of using the self-example training set, the training must be performed on-line. Using the multi-threading programming technique can improve the efficiency of the training.

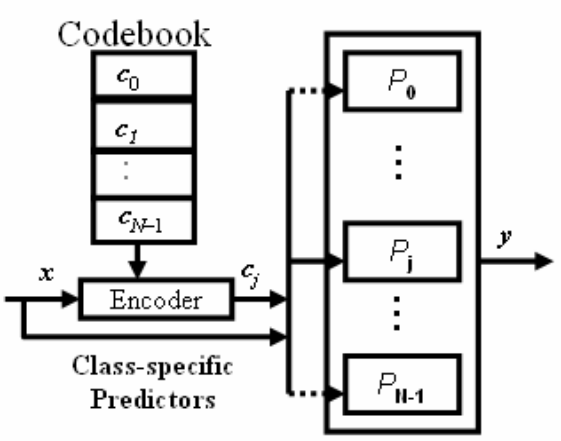


Figure 1 A block diagram of our example-based image super-resolution algorithm, which is composed of a content-based encoder in the form of a vector quantizer, and a group of class-specific predictors to infer the high-frequency details.

2.4. High-Resolution Image Reconstruction

Having trained the content-based encoder and the class-specific predictors, the HR version of a LR image can be constructed. The input LR image is first magnified using the bilinear interpolation to form an initial estimation of its HR version, denoted as $\hat{I}_0$. The high-frequency layer L_0 is estimated using one of the N class-specific predictors, and is then added to the initial estimated image to construct a HR image with high visual quality, i.e.

$$I_0 = \hat{I}_0 + L_0. \quad (4)$$

Each 4×4 block B_h in the HR image has a corresponding 4×4 LR block B_l in the difference image L_1 of the input LR image. The central 2×2 patch of B_l is the low-resolution version of B_h. In our implementation, in order to handle those blocks at the boundary of L_1, all of the pixels at the boundary are extended and duplicated by one pixel. The block B_l is then encoded and classified to one of the categories, and the corresponding class-specific classifier is employed to infer the high-frequency information about B_h. Note that the reconstructed HR blocks are demeaned and have unit variance, so they are transformed to have the original means and variances. In our algorithm, the HR block B_h is shifted by a step of 2 in the horizontal and vertical directions, and the corresponding LR block B_l is shifted by a step of 1 accordingly. At each position of the blocks, the high-frequency information is predicted using an

appropriate class-specific predictor. Then, the overlapped high-frequency information is averaged to produce an estimation of the high-frequency layer.

Finally, the high-frequency layer is added to the initial estimated image, as in (4), and a LR constraint is also applied to the resulting image. We assume that the reconstructed HR image can produce the input LR image by smoothing and sub-sampling. The image I_0 is blurred and down-sampled to form the LR image I_1. The average of a $z \times z$ block in I_0 will correspond to a single pixel in I_1. Suppose that the average value in I_0 and the corresponding single pixel values in I_1 are p_i and q_i, respectively. Then, the error is computed as follows:

$$e_i = q_i - p_i . \quad (5)$$

This error value is added to each pixel in the $z \times z$ block to reconstruct the final HR image.

3. EXPERIMENTS AND DISCUSSIONS

We will evaluate the performance of our proposed algorithm with the use of two different training sets that mentioned in Section 2.1. Two different types of images will be considered in our experiments: face images and natural-scene images. For each type of training set, the optimal number of categories for content-based encoding determined based on experiments is used, viz. 28 for face images and 68 for natural images. The visual qualities and the computational complexities of our algorithm in combination with each of the different training sets will be measured.

For the self-example training set, the images themselves are used for training as well as for testing. For domain-specific applications, we consider the super-resolution of face images. Therefore, a number of face images and natural-scene images are used in the experiments. For the face images, the ORL database [19] is employed, which contains 40 distinct subjects, and each subject has 10 different images of size 92×112 pixels. In addition, to evaluate the performance of our algorithm for different types of images, a set of natural-scene images is used. The images have very different appearances to each other. Concerning the domain-specific training set and the general-purpose training set, 50 % of the face images and the natural-scene images, respectively, is selected for training, while the remainder will be used for testing. Figure 2 shows some training images in the ORL database.

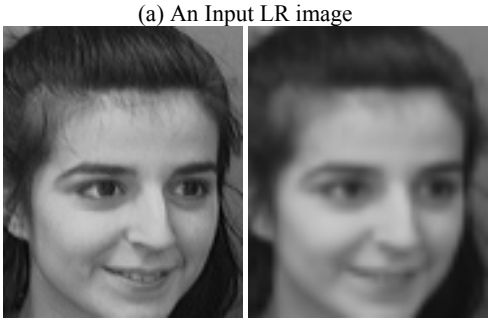
Figure 3 and Figure 4 illustrate some of the images using different image super-resolution algorithms. The first one, i.e. Fig. 2(a) and Fig. 3(a) show the input LR images of size 46×56 for the face images and 128×128 for the natural-scene images, which are down-sampled from the original HR images shown in Fig. 2(b) and Fig. 3(b). The images in Fig 2(c) and Fig. 3(c) are the results generated by bilinear (#1) interpolation. The results shown in Fig. 2(d) and Fig. 3(d) are based on Chen [14] (#2), which is a "searching and pasting" method. We can see that the visual quality of these images is improved to a certain extent when compared to those achieved by bilinear interpolation. The mouth region and the eye regions contain more high-frequency details. However, the unmatched patches for these regions will greatly degrade the reconstruction quality. The last images, i.e. Fig. 2 (e) and Fig. 3(e), are produced using our algorithm (#3) with the self-example training set. Because of the use of class-specific predictors in our algorithm, the unmatched problem can be avoided, and the image quality is improved.

The PSNR and MSE are objective measurements of image quality, and they need not be consistent with subjective human visual perception. Table 1 tabulates the average PSNR, MSE, and runtime of the different algorithms. The results obtained using our algorithms are based on the use of the optimal number of levels. We can see that, with the self-example training set, our algorithm can achieve a smaller MSE, and therefore, a higher PSNR as compared to the other two algorithms. The average PSNR, MSE, and runtime of our algorithm using the Set B for face images is also tabulated in the Table 1.

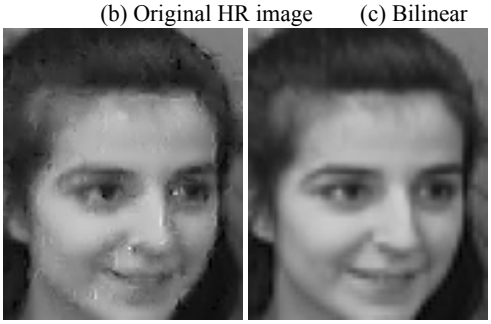
The experiments were executed on an Inter® Core™ 2 CPU 6600 @2.40GHz with 2 GB RAM system. Our algorithm can achieve a shorter runtime than other example-based algorithms for two reasons: the self-example training set is of a small size, with its content correlated; and the class-specific predictors can be designed in parallel by using multi-thread programming. As for the "searching and pasting" method, it requires searching a huge training set for each block of an input image, so it is more computationally intensive.

Figure 2 Some training images in the ORL database

(a) An Input LR image



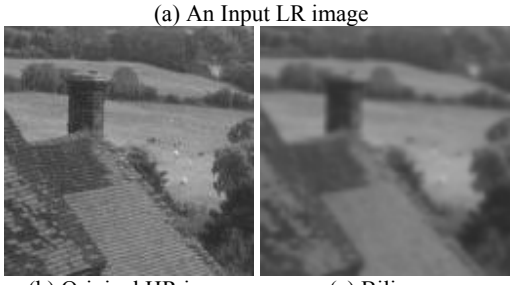
(b) Original HR image (c) Bilinear



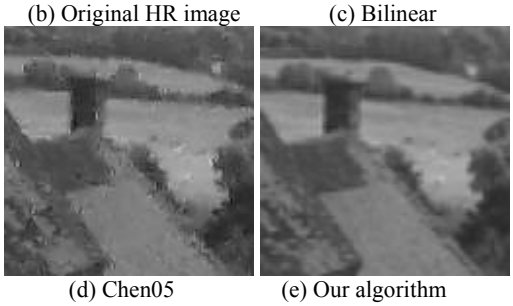
(d) Chen05 (e) Our algorithm

Figure 3 Experimental results based on the ORL database.

(a) An Input LR image



(b) Original HR image (c) Bilinear



(d) Chen05 (e) Our algorithm

Figure 4 Experimental results based on a natural-scene image.

Table. 1 Performance of different algorithms.

Images/Algori.		#1	#2	#3	
				Set A	Set B
Face Image	PSNR (dB)	27.75	26.92	29.78	30.00
	MSE	118.17	140.7	73.69	70.34
	TIME (s)	0.01	12.59	0.35	5.752
Natural Image	PSNR (dB)	30.70	28.61	32.25	——
	MSE	64.64	90.18	39.84	——
	TIME (s)	0.015	82.02	7.203	——

4. CONCLUTIONS

The example-based approach is a promising way to solve the image super-resolution problem, which can provide the high-frequency contents of a reconstructed HR image by learning. However, most of the existing algorithms interpret the "learning" as just a kind of "searching" the best-matched LR patch, and then "pasting" the corresponding HR component. In our algorithm, we improve the learning by using a set of class-specific predictors, where the prior high-resolution information is stored as the weights of the predictors. The content of a training set is more important than its size. In order to exploit the efficiency and effectiveness of training sets, a self-example set, a domain-specific training set, and a combined set have each been investigated in experiments. Experimental results show that our algorithm can achieve an excellent performance in terms of both quality and computational complexity.

5. ACKNOWLEDGEMENT

This work was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 5199/06E), and by the National Nature Science Foundation of China (60472036, 60431020, 60402036, 60772069, 60532040), the Natural Science Foundation of Beijing (No. 4062006), and the Beijing Novel Program (2005B08).

6. REFERENCES

[1] S. C. Park, M. K. Park and M. G. Kang, "Super-resolution image reconstruction: A technical overview,"

IEEE Signal Processing Magazine, vol. 5, pp.21-36, 2003.

[2] H. A. Aly, E. Dubois, "Image up-sampling using total-variation regularization with a new observation mode," IEEE Trans. on Image Processing, vol.14 no.10, pp. 1647-1659, 2005.

[3] S. Farsiu, M. D. Robinson. and M. Elad, *et al*. "Fast and robust multiframe super resolution," IEEE Trans. on Image Processing, vol. 14, no. 10, pp. 1327-1343, 2004.

[4] H. He, L. P. Kondi, "An image super-resolution algorithm for different error levels per frames," IEEE Trans. on Image Processing, vol. 15, no. 3, pp. 592-603, 2006.

[5] G. K. Chantas, N. P. Galatsanos and N. A. Woods, "Super-resolution based on fast registration and maximum a posteriori reconstruction," IEEE Trans. on Image Processing, vol. 16, no. 7, pp. 1821-1830, 2007.

[6] S. Baker, T. Kanade, "Limits on super-resolution and how to break them," IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2, pp. 372-379, 2000.

[7] S. Baker, T. Kanade, "Limits on super-resolution and how to break them," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, pp. 1167-1183, 2002.

[8] W. T. Freeman, E. C. Pasztor, "Learning low-level vision," International Journal of Computer Vision, vol. 40, no. 1, pp. 25-47, 2000.

[9] W. T. Freeman, T. R. Jones and E. C. Pasztor, "Example-based super-resolution. IEEE Computer Graphics and Applications," vol. 22. no. 2, pp. 56-65, 2002.

[10] Q. Wang, X. Tang and H. Shum, "Patch based blind image super resolution," In: Proc. of the Tenth IEEE International Conf. on Computer Vision, Beijing, China, 2005, Oct.

[11] T. A. Stephenson, T. Chen, "Adaptive markov random fields for example-based super-resolution of faces," Journal on Applied Signal Processing, vol. 2006, pp. 1-11, 2006.

[12] G. Qiu, "Interresolution look-up table for improved spatial magnification of image," Journal of Visual Communication and Image Representation, vol. 11, pp. 360-373, 2000.

[13] M. Elad, D. Datsenko, "Example-based regularization deployed to super-resolution reconstruction of single image," The Computer Journal Advance Access published online on April, 20, 2007.

[14] M. Chen, G. Qiu and K. M. Lam, "Example selective and order independent learning-based image super-resolution," In: Proc. of 2005 International Symposium on Intelligent Signal Processing and Communication Systems, pp. 77-80, 2005.

[15] X. Zhang, K. M. Lam and L. Shen, "Image magnification based on adaptive MRF model parameter estimation," In Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems, Hong Kong, 2005.

[16] M. Ebrahimi, E. R. Vrscay, „Solving the inverse problem of image zooming using 'self-examples'," In: M. S. Kamel, A. C. Campilho (Eds), ICIAR Lecture Notes in Computer Science, Springer, vol. 4633, pp. 117-130, 2007.

[17] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. on Communications, vol. 28, no. 1, pp. 84-95, 1980.

[18] G. Qiu, "A progressively predictive image pyramid for efficient lossless for coding," IEEE Trans. on Image Processing, vol. 8, no. 1, pp. 109-115, 1999.

[19] F. Samaria, A. Harter, "Parameterisation of a stochastic model for human face identification," In: 2nd IEEE Workshop on Applications of Computer Vision, Sarasota, Florida. Dec. 1994.