ENLARGEMENT OF DIGITAL IMAGES ON LAPLACIAN PYRAMID BY USING NEURAL NETWORK

Daisuke SEKIWA and Akira TAGUCHI

Department of Electrical and Electronic Engineering, Musashi Institute of Technology
e-mail taguchi@ipc.musashi-tech.ac.jp

ABSTRACT

A neural network for edge-preserving image interpolation is introduced, which is based on the non-linear procedure which is presented by Greenspan [2]. Simulation results show the superior performances of the proposed approach with respect to other interpolation techniques.

1 Introduction

Enlargement of images corresponds to narrow Nyquist interval in digital images. Usually, enlargement of images are given by interpolation by using sinc function based on sampling theorem. Generally, there are infinite frequency components in natural images. Sampling images, there are lost high frequency components in digital images. Thus, we need for creating images with higher resolution than the sampling rate would allow. However, enlarged images by using sinc function don't have high frequency components which were lost by sampling. Greenspan had presented the procedure for predicting an unknown higher resolution image, which utilizes Laplacian pyramid image representation [2]. This paper presents a novel enlargement method by using a neural network(NN) which is extended Greenspan's non-linear procedure. Results are presented depicting the visual enhancement of several images. Several simulation results are presented the proposed method is superior to the Greenspan's method. Furthermore, we show the robustness of the proposed method.

2 The Enlargement by Non-linear Method

Gaussian and Laplacian pyramids, as described by Burt [1], are utilized. The Gaussian pyramid consists of lowpass filtered versions of the input image as G_0, with each stage image as G_{n+1} of the pyramid achieved by Gaussian filtering of the previous stage one as G_n and corresponding sub-sampling of the filtered output. The Laplacian pyramid consists of bandpass filtered versions of the input image, with each stage of the pyramid constructed by the subtraction of two corresponding adjacent levels of the Gaussian pyramid. A recursive procedure allows for the creation of the pyramids, as follows:

$$\bar{G}_{n+1} = W * G_n, \tag{1}$$

$$G_{n+1} = \text{down sampling } \bar{G}_{n+1}, \tag{2}$$

$$L_n = G_n - \mathit{EXPAND}(G_{n+1}), \tag{3}$$

$$\mathit{EXPAND}(G_{n+1}) = 4 \times (W * G^0_{n+1}), \tag{4}$$

where "$*$" show the convolution operation and G^0_{n+1} whose size is the same with G_n is the interpolated zero of G_{n+1}. Generally, the weighting function W is Gaussian in shape and normalized to have the sum of its coefficients equal to 1. The value for a 5-sample separable filter are $(1/16, 1/4, 3/8, 1/4, 1/16)$. The higher level image G_{-1} than the original digital image G_0 is got $G_{-1} = L_{-1} + \mathit{EXPAND}(G_0)$. However, L_{-1} is unknown and it is necessary to predict a higher frequency component L_{-1}.

An edge can be characterized by zero crossings (ZC) in the Laplacian image. Main edges are observed across different resolution images. Utilizing the character, Greenspan has presented a procedure for predicting L_{-1} by non-linear method using L_0. The algorithm includes the following 3 steps:

Step 1 We extract the high frequency components L_0 of an image G_0 and create $\mathit{EXPAND}(L_0)$ from the L_0

$$L_0 = G_0 - EXPAND(G_1). \tag{5}$$

Step 2 We obtain $\hat{L}_{-1}$ by thresholding as follows.

$$\begin{aligned}\hat{L}_{-1} &= \mathit{BOUND}\{\mathit{EXPAND}(L_0)\}\\ &= \alpha \times \begin{cases} T & \text{if } \mathit{EXPAND}(L_0) \geq T \\ \mathit{EXPAND}(L_0) & \text{if } -T < \mathit{EXPAND}(L_0) < T \\ -T & \text{if } \mathit{EXPAND}(L_0) \leq -T \end{cases}\end{aligned} \tag{6}$$

The image $\hat{L}_{-1}$, which is obtained by non-linear operation, has high frequency components than the sampling rate.

Step 3 We create $\tilde{L}_{-1}$ by Laplacian filtering the $\hat{L}_{-1}$ to reshape the new transients so they have the desired spatial frequency components

$$\tilde{L}_{-1} = \hat{L}_{-1} - W * \hat{L}_{-1}. \qquad (7)$$

Since, we guess that a higher frequency component L_{-1} preserves the shape and the phase of L_0, and L_{-1} is approximated by $\tilde{L}_{-1}$ which is obtained by the above approach.

In the following, the experiment with one dimensional step signal is discussed. Fig.1a illustrates the ideal signal G_{-1}, the expanded signal $EXPAND(G_0)$, and the predicted signal by Greenspan's method (step 1–3). Fig.1b illustrates the Laplacian components L_{-1}, $EXPAND(L_0)$, and $\tilde{L}_{-1}$ similarly. We can verify that $\tilde{L}_{-1}$ has the same ZC with L_1 and sharpens the waveform of $EXPAND(L_0)$ from Fig.1b. However, the value of "T" and "a" in Eq. 6 are different in each image and obtained empirically.

3 Enlargement Neural Network

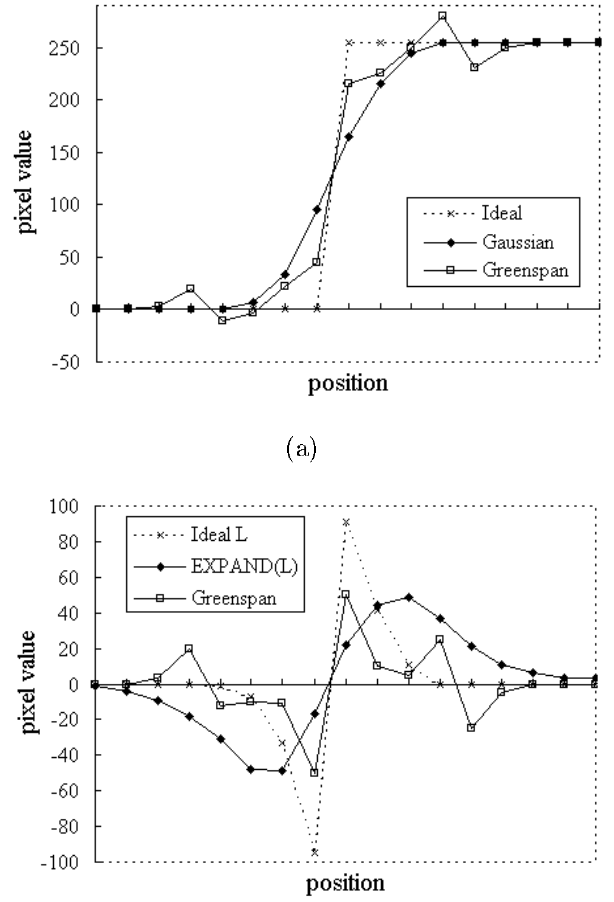
We can express above Greenspan's non-linear scheme as partly connected feedforward three layer neural network(NN).

$$\begin{aligned} L_n &= G_n - EXPAND(G_{n+1}) \\ &\approx G_n - W * G_n \qquad (8) \\ EXPAND(L_n) &\approx EXPAND(G_n) - W * EXPAND(G_n) \\ &= (1 - W) * EXPAND(G_n) \\ &= 4 \times (1 - W) * W * G_n^0 \qquad (9) \end{aligned}$$

Step 1 of Greenspan's method (*i.e.*, Eq. 9) is realized by between input layer and hidden layer. Step 2 of Greenspan's method is carried out the non-linear function of nodes in hidden layer. The high frequency components $\tilde{L}_n$ is extracted between hidden layer and output layer (*i.e.*, $\tilde{L}_n = \hat{L}_n - W * \hat{L}_n$). Then, $\tilde{L}_n$ is obtained from the NN's output layer.

There are linear nodes in the input and the output layer. Since the weight function W is 5×5 window, we need 5×5 nodes in the hidden layer for getting one output from the output layer. We also need twice Gaussian filtering in Eq. 9. We accordingly set to 13×13 the number of nodes in the input layer. Thus, the three layer NN consists (input nodes, hidden nodes, output nodes) = (13×13, 5×5, 1). We denote the relations among input layer input "I_j" , hidden layer output "H_j" and output layer output "O".

$$U_j = \sum_{i=1}^{169} w_{ji}^1 I_i \qquad (10)$$



(b)

Figure 1: Ideal and enlarged signals, in case of (a) ideal G_{-1}, $EXPAND(G_0)$, and Greenspan's method; (b) ideal L_{-1}, $EXPAND(L_0)$, and Greenspan's method $\tilde{L}_{-1}$.

$$H_j = f(U_j) \qquad (11)$$

$$O = \sum_{j=1}^{25} w_j^2 H_j \qquad (12)$$

Where "$f(x)$" is presented by

$$f(x) = BOUND(x). \qquad (13)$$

The enlargement NN is trained in order to minimize the mean square error (MSE) between teaching signal L_n and NN output O by

$$\text{minimize:} J(\mathbf{w}^1, \mathbf{w}^2) = E[\{O - L_n\}^2].$$

$\mathbf{w}^1$ and $\mathbf{w}^2$ are the weight vectors which consist of w_{ji}^1

and w_j^2. Initial value of weights are set to an equal value with Greenspan's method.

4 Simulation Results

4.1 One Dimension Signal

Here, we apply the enlargement NN to the same signal as Fig.1. Fig.2 illustrates the ideal signal, the cubic convolution signal, the expanded signal by Greenspan's method, and the expanded signal the proposed NN. We would recognize that the enlargement NN produces the Laplacian component which is the most close to the ideal one from Fig.2b. The enlargement NN can naturally restructures the sharper edge than Greenspan's method (Fig.2a). This shows that the enlargement NN is the network which achieves to restore the lost high frequency components at downsampling and gets the optimum condition of Greenspan's method. It is good result of cubic convolution, but it does not have enough high frequency components than even Greenspan's method.

4.2 Two Dimension Signal

We apply the enlargement NN to enlarge an image. We show the robustness of the proposed NN following 3 points of view.

1. The enlargement NN is constructed with independence from levels of a Laplacian pyramid representation.
2. The enlargement NN is constructed with independence from different kinds of images.
3. The enlargement NN is constructed with independence from reduction procedures, which are generally unknown, to get the low resolution image from the high resolution image. Here we assume that it is Gaussian filtering then downsampling as the reduction procedure.

We now examine the point 1. We calculate six enlargement NNs with "Lena" and "Boat" as follows:

NN-1L: input image: Lena G_1^0, teaching image:L_0

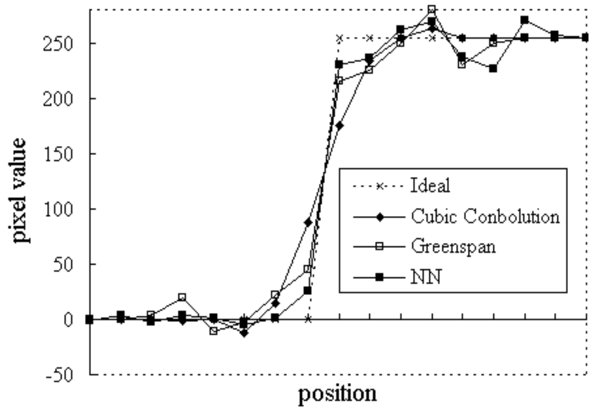
NN-2L: input image: Lena G_2^0, teaching image:L_1

NN-3L: input image: Lena G_3^0, teaching image:L_2

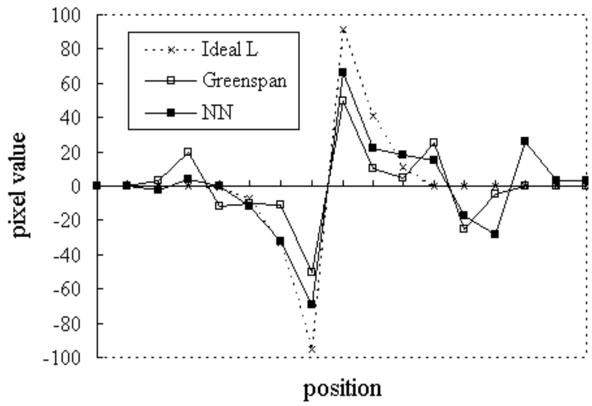
NN-1B: input image: Boat G_1^0, teaching image:L_0

NN-2B: input image: Boat G_2^0, teaching image:L_1

NN-3B: input image: Boat G_3^0, teaching image:L_2



(a)



(b)

Figure 2: Ideal and enlarged signals, in case of (a) ideal G_{-1}, $EXPAND(G_0)$, Greenspan, and the NN; (b) ideal L_{-1}, Greenspan's $\tilde{L}_{-1}$, the NN.

We enlarge images (from G_3 to G_2, from G_2 to G_1, and from G_1 to G_0 of "Lena" and "Boat") by six enlargement NNs trained different levels. Table 1 shows the resulting MSEs between the desired image and the enlargement images. PSNR in Table 1 is given by

$$PSNR = 10 \cdot \log \frac{(255)^2}{MSE} \quad [dB]. \tag{14}$$

It is seen that the three enlargement NNs give similar results in the MSE, and hence we recognize that the enlargement NNs is constructed with independence from levels.

In the following, we examine the point 2. Here, we enlarge the images from G_1 to G_0 of "Lena", "Girl", "Boat", "Light House", and "Building" by using the

NN-1L. Table 2 shows the resulting MSEs by the NN-1L, Greenspan's method and cubic convolution interpolation(CCI). Excellent enlarged images are obtained by the NN-1L at all examples. Comparing the NNs with CCI, the NNs show better results of 1.2dB about "Light House" and 2dB about the others. We would like to show that the enlargement NN is constructed with independence from different kinds of images.

Also we show the subjective evaluation in Fig.3 which show the result of enlarged images of "Boat" by the NN-1L. Fig.3a shows the desired image. From Fig.3, the proposed NN recover the high-frequency components which were lost by sampling.

At last, we examine the point 3. In this paper, we assume that there is Gaussian filtering and downsampling as the reduction procedure. However, we can not generally know how to be reduced images, then we show the NN's characteristic by reduction processes. Here, we use three kind of it, which is a) Gaussian filtering and downsampling (GF), b) 3×3 mean value filtering and down sampling (AF), and c) cubic convolution filtering and downsampling (CC). We expand reduced images of "Boat" and "Light House", which have different reduction procedure, by the NN-1L, Greenspan's method, and CCI. Table 3 shows the resulting MSEs by the NN-1L, Greenspan's method and CCI. Reducing by GF and AF, the NN-1L has the best results than the other expand procedures. However, reducing by CC, Greenspan's method and CCI has desired results than the NN-1L numerically. To be obviously the causes, we show the subjective evaluation of expanded images by the NN-1L and CCI in Fig.4. In Fig.4, we recognize that the edge and detail signals are enhanced by the NN-1L. Cubic convolution function has superior frequency characteristic near Nyquist frequency to Gaussian filter, then we consider that the NN-1L, which is taught by Gaussian filtered image, enhances high frequency components more than we expect. However, the fine lines as wire at the enlarged image by the NN-1L are more sharper than one of CCI, and thereby we would recognize the increase of resolution by the NN-1L with independence from reduction procedures.

5 Conclution

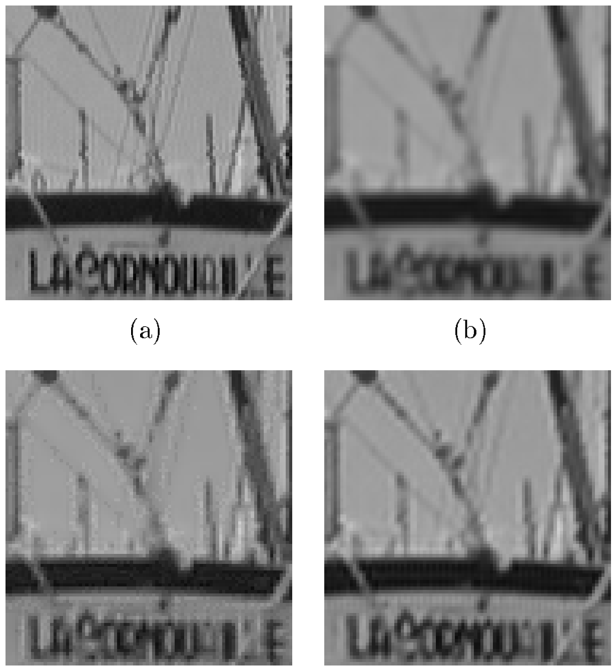
This paper presents a procedure for predicting a higher-frequency component than the sampling rate would allow. This NN is general three layer NN, and contains Greenspan's algorithm which predicts higher level Laplacian image by using the Laplacian pyramid. Excellent enlarged images are obtained by the enlargement NN. Furthermore, the characteristics of the enlargement NN are not depended on types and levels of

Table 1: The compare the enlargement NNs constructed by different levels. (MSE (PSNR[dB]))

Lena	G_1 to G_0	G_2 to G_1	G_3 to G_2
NN-1L	75.9(29.3)	38.4(32.3)	46.6(31.4)
NN-2L	76.3(29.3)	35.0(32.7)	44.9(31.6)
NN-3L	79.3(29.1)	39.2(32.2)	40.9(32.0)
Boat	G_1 to G_0	G_2 to G_1	G_3 to G_2
NN-1B	63.8(30.1)	47.1(31.4)	54.0(30.8)
NN-2B	67.5(29.8)	43.9(31.7)	51.6(31.0)
NN-3B	74.0(29.4)	50.5(31.1)	49.3(31.2)

Table 2: Enlargement results of "Lena", "Boat", "Girl", "Light House", and "Building" from G_1 to G_0. (MSE (PSNR[dB]))

enlarge\image	Lena	Boat	Girl	Light House	Building
CCI	115.7(27.5)	102.5(28.0)	44.9(31.6)	286.4(23.6)	96.1(28.3)
Greenspan	106.4(27.9)	90.0(28.6)	46.3(31.5)	254.5(24.1)	89.7(28.6)
NN-1L	75.9(29.3)	65.3(30.0)	28.1(33.6)	217.8(24.8)	55.8(30.7)



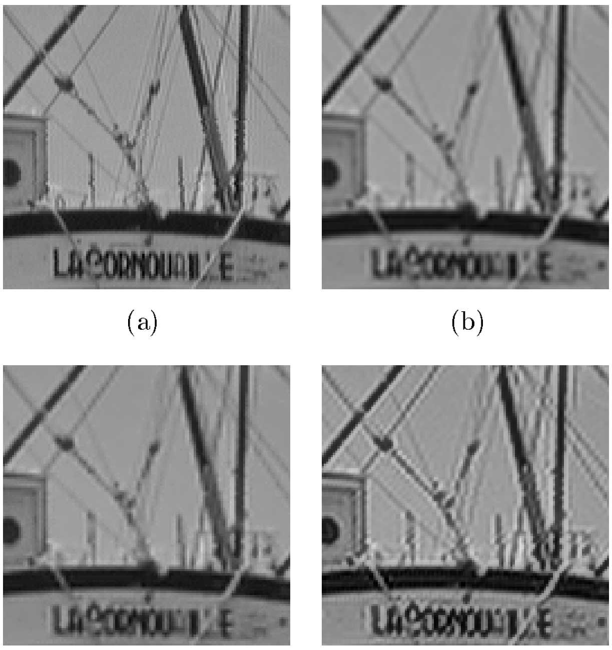
(a) (b) (c) (d)

Figure 3: Ideal and enlarged images; (a)original, (b)CCI, (c)Greenspan, (d)NN.

training image so much.

Table 3: Enlargement results from various reduced images. (MSE (PSNR[dB]))

Boat	GF	AF	CC
CCI	102.5(28.0)	89.2(28.6)	74.3(29.4)
Greenspan	90.0(28.6)	84.4(28.9)	70.9(29.6)
NN-1L	65.3(30.0)	70.2(29.7)	129.3(27.0)
Light H.	GF	AF	CC
CCI	286.4(23.6)	265.3(23.9)	228.1(24.5)
Greenspan	254.5(24.1)	248.2(24.2)	226.9(24.6)
NN-1L	217.8(24.8)	225.4(24.6)	297.5(23.4)



(a) (b) (c) (d)

Figure 4: Ideal and enlarged images; (a)original, (b)CCI, (c)Greenspan, (d)NN.

6 References

[1] P. J. Burt and E. A. Adelson, "The Laplacian Pyramid as a compact image code." IEEE Trans. Commun., Vol.COM–31, pp.532–540, 1983.

[2] H. Greenspan and C. H. Anderson, "Image enhancement by non-linear extrapolation in frequency space." SPIE Vol.2182 Image and Video Processing II, pp.2–13, (1994)