

SOFM과 LAM을 이용한 영상 보간에 관한 연구

장동언, 정태상
중앙대학교 제어계측공학과

A Study on Image Interpolation Using SOFM and LAM

Dong-Eon Chang and Tae-Sang Chung
Department of Control and Instrumentation Engineering Chung-Ang University

Abstract - When resampling an image to a new set of coordinates, there is often a noticeable loss in image quality. The interpolation kernel determines the quality of interpolation. In this paper, We think two interpolation methods: cubic-spline method, neural net method, at first study given interpolation method using spline and then present new interpolation method using SOFM and LAM(neural net method), finally compare the performance of several interpolation methods including replication, bilinear, spline and new methods.

1. 서 론

보간(Interpolation)이란 주어진 표본(Samples)사이에 주어지지 않은 함수의 값을 찾아내는 과정이다. 영상 보간은 정지 영상이나 동영상에서 영상 압축, 영상 확대 등 여러 가지의 용용분야에 중요한 역할을 하기에 중요성이 증대되고 있다.

일반적인 영상은 2차원 신호로서 이웃한 영상 신호 사이에는 상관관계(Correlation)가 많고 공간상으로 저주파 성분이 우세하다[1].

복사 보간법(Replication method)은 보간할 영상을 가장 가까운 픽셀의 값을 그대로 취하는 가장 단순한 방법이다. 선형 보간법(Bilinear method)은 이웃한 두 픽셀값의 선형치를 보간할 값으로 취하는 방법이다.

이 논문에서는 먼저 1차원 필터로서 분리가능한 2차원 필터를 구성하여 영상을 복원하는 삼차 스플라인(Cubic Spline)보간법에 관해서 학습하고 삼차 보간법에서 커널(kernel) 특성의 변화를 조정가능하게 하는 변수를 확인하고, 그 변수를 하이퍼볼릭 탄젠트(Hyperbolic tangent) 모서리 모델에 관해 MSE(Mean Square error) 또는 MAE(Maximum Absolute Error)를 최소화하는 값을 찾겠다.

영상의 이웃 신호의 상관관계를 이용하고 영상의 모서리 특성을 적용하기 위해, 학습 영상에 대해 SOFM(Self-Organized Feature Map) 신경망을 이용하여 Codebook을 구성한다. 각각의 Codeword에 대해 LAM(Linear Associative Memories)을 이용하여 저해상도 영상과 고해상도 영상의 최적 관계를 결정짓는다. 영상 보간은 보간할 영상을 블락단위로 구성하여 각각에 해당되는 Codeword를 찾고 이에 해당하는 고해상 블락을 LAM을 통하여 대치하면 된다.

끝으로 우리가 구성한 보간방법을 일반적으로 이용하는 Bilinear, Spline 보간방법에 대하여 SNR측면에서 성능을 비교하겠다.

2. 본 론

2.1 Spline을 이용한 보간 방법에 관한 비교

삼차 스플라인 보간법은 1차원적 방향에 있어서 가장 가까운 이웃한 네 픽셀값을 이용하여 각 픽셀 값을 지나고 1차 미분값이 같으며 중심에 대하여 대칭으로 보간하는 방법이다. 아래에서는 기존의 스플라인 방법을 학습하고 모서리 정보에 관하여 최적인 스플라인 계수를 찾겠다.

2.1.1 스플라인을 이용한 기존의 보간 방법

3차 B-스플라인 방법[2]은 수식(1)과 같다.

$$f(x) = \frac{x^3}{2} - x^2 + 4/6 \quad (0, 1)$$

$$f(x) = -x^{3/6} + x^2 - 2x + 8/6 \quad (1, 2) \quad (1)$$

일반화된 3차 스플라인 보간방법[3]은 수식(2)와 같다.

$$f(x) = a_{30}x^3 + a_{20}x^2 + a_{10}x + a_{00} \quad (0, 1)$$

$$f(x) = a_{31}x^3 + a_{21}x^2 + a_{11}x + a_{01} \quad (1, 2) \quad (2)$$

수식 (2)에서 커널의 특성을 규정하기 위하여 x 가 0인 점에서의 함수값은 1이고 x 가 1, 2인 점에서는 함수값이 0을 가져야 한다. 그리고 x 의 값이 0.1에서는 연속이어야 하며, 0.2에서의 함수 기울기는 0이야 하며 x 가 1에 스는 함수의 좌미분값과 우미분값이 일치하여야 한다. 이것을 조건으로 수식(2)를 정리하면 수식(3)과 같다.

$$f(x) = (a+2)x^3 - (a+3)x^2 + 1 \quad (0, 1)$$

$$f(x) = ax^3 - 5ax^2 + 8ax - 4a \quad (1, 2) \quad (3)$$

위의 수식(3)에서 a 를 $-1/2$ 로 하면 주파수 영역에서 저주파성분은 평행하며 고주파 성분은 차단주파수 이상에서는 급격히 0으로 간다. 그리고 Tayler 시리즈 근사법에서 2차 다항식까지 일치한다. 또한 x 가 1인 곳에서의 기울기를 sinc함수의 기울기인 1에 일치시키면 a 가 -1 이 된다. 그리고 x 가 1인 곳에서의 2차 좌미분, 우미분값이 일치하는 a 를 찾으면 $-3/4$ 의 값을 갖는다.

일반화된 3차 스플라인 방법은 3차 B-스플라인 방법보다 주파수 영역에서 좋은 성능을 보인다.

2.1.2 모서리 정보를 이용한 스플라인 보간 방법

위의 수식(3)을 일반적인 모서리 모델인 하이퍼볼릭 탄젠트 모델에 적용하여 오차가 최소화하는 측면에서 찾아보면 최대 절대오차를 최소화하는 값은 $a = -0.578$ 이고 제곱평균오차를 최소화하는 값은 $a = -0.609$ 이다. 이값은 기존의 일반적인 스플라인 방법에서의 $a = -1/2$ 값에 근접하는 값으로서 비슷한 주파수 영역에서의 성능을 보인다.

2.2 SOFM과 Associative Memory를 이용한 보간 방법[4]

이 영상 보간 방법에서는 영상에서 상관관계를 이용하기 위하여 보간할려는 영상의 국소부분에 대하여 각각의 다른 보간 커널을 사용한다. 즉 보간 할 국소 부분의 영상 특성에 따라서 해당되는 최적의 커널을 선택하고 각 선택된 커널에 대하여 각각의 픽셀에 2차원적인 보간을 한다.

여기서 각각의 커널에 대한 codebook를 만들기 위하여 SOFM방법을 사용하며 각각의 codeword에 대한

최적의 대응을 위하여 LAM을 이용한다.

고해상도 근본 영상을 X_{high} 라 하고, 근본 영상에 대응하는 저해상도 영상을 X_{low} 라 하겠다. 그리고 신경망을 통하여 보간된 고해상도 영상을 \hat{X}_h 로 표시하겠다. 그리고 G_x, G_y 는 X 축, Y 축 방향으로의 감쇄 비율을 나타내며 결국, 확대될 비율이 될 것이다.

먼저 각각의 X_{high} 를 감쇄 기법을 이용하여 G_x, G_y 만큼 다운샘플된 X_{low} 를 구한다. X_{low} 를 국소화 시키기 위하여 H_x, H_y 크기의 단위로 구분하여 벡터화($V(\cdot)$)시키고 평균값으로 펜다. 벡터화는 2차원 구조를 열단위로 차근차근 쌓으면 된다. 벡터화 된 영상을 전처리를 사용하여 입력으로 사용한다.

그림1은 신경망을 이용한 영상보간의 전체적인 구성도를 나타낸다. 그림1(a)는 신경망의 학습과정을 나타내며, 그림1(b)는 저해상도 영상의 보간에 관한 과정을 나타낸다.

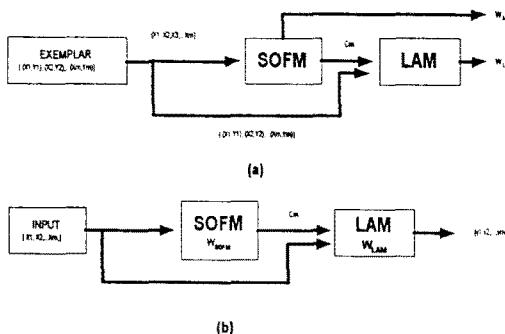


그림 1

신경망을 이용한 영상보간의 구성도

2.2.1 SOFM을 이용한 Clustering

X_{low} 의 각각의 국소부분은 특성의 유사도에 따라 clustering된다. clustering는 Kohonen의 SOFM을 이용하여 이루어 진다. SOFM은 기본적으로 입력특성에 대하여 기하학적 순서로 벡터화를 만들어 준다. X_{low} 의 i번째 입력을 $X_{low,i}$ 라 하고 SOFM은 전체 M개의 마디를 가지도록 구성하고 각각의 마디를 F_m 으로 표시하자. 여기서 $m = 1, 2, \dots, M$ 이다.

그러면 각각의 $X_{low,i}$ 은 수식 (4)에 의해 cluster C_m 에 해당된다.

$$C_m = \{ X_{low,i} : \|F_m - V(X_{low,i})\|_2 < \|F_n - V(X_{low,i})\|_2 \} \quad (4)$$

여기서 $m = 1, 2, \dots, M$ 이고 $m \neq n$ 이다.

즉 영상에서 국소 특성이 Euclidean 거리에서 특성마다 F_m 에 가장 가까우면 cluster C_m 의 구성원이 된다.

2.2.2 저해상도 영상과 고해상도 영상의 대응(Association) 확립

학습 영상에서 SOFM에 의해 구성된 각각의 C_m 의 모든 구성원에 대하여 X_{low} 의 $H_x \times H_y$ 단위의 국소 영상과 이에 대응하는 X_{high} 의 $(2G_x-1) \times (2G_y-1)$ 을 MSE(Mean Square Error)측면에서 최적으로 대응하는 LAM을 구성하는 과정이다.

그림2에서 저 해상도 영상($H_x \times H_y$)은 검은 점으로 표시되고 고해상도 영상($(2G_x-1) \times (2G_y-1)$) 즉 보간될 영상은 \otimes 으로 표시된다. 그림1은 H_x, H_y 를 3으로 하였으며, G_x, G_y 는 2로 하였다.

저해상도의 영상에 대하여 고해상도의 영상을 대응시킴으로서 LAM을 구성한다. 여기서 LAM은 근본적으로 연

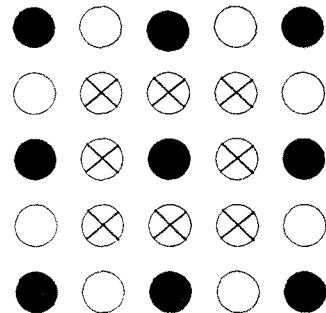


그림 2 저해상도 영상과 고해상도 영상의 대응

속적인 입력 영역을 연속적인 출력 영역에 함수적인 대응이다. 이러한 대응은 학습영상에서 같은 cluster에 속하는 구성원들의 유한한 갯수의 입력들(저해상도 픽셀들)과 출력들(고해상도 픽셀들)을 대응하여 LAM의 Weight를 학습시키는 과정이다.

각각의 C_m 에 대하여 입력집단을 V 라 하면 이 행렬의 각각의 열 V_r 은 저해상도 구성원을 벡터화 과정을 통하여 열로 만든 것이다. 같은 방법으로 원하는 출력집단을 D 로 표현하고 V 를 입력으로 할 때 LAM을 통하여 나온 실제 출력을 Y 로 하자.

LAM의 입출력 관계는 (5)과 같다.

$$Y_r = W V_r + B \quad (5)$$

W 는 LAM의 Weight 행렬이고 B 는 bias 행렬이다. 각각의 C_m 에 대하여 W_m 과 B_m 을 구해서 저해상도의 국소 특성에 맞는 커널을 제작해야 한다.

그리고 최적의 LAM은 제곱 평균 오차를 최소화하는 것이다. 즉 식(6)와 같이 오차(E)를 최소화 한다.

$$E = \sum (D_r - Y_r)^2 \quad (6)$$

최적의 W_m 과 B_m 은 LMS알고리즘을 제거적으로 사용하여 수식(7)과 같이 구할 수 있다.

$$[W_m | B_m](t+1) = [W_m | B_m](t) + \mu(D - Y)[V^T | U] \quad (7)$$

여기서 A^T 는 행렬 A 의 전치행렬을 의미하고, μ 는 학습율을 뜻하며 U 는 원소가 모두 1인 열벡터이다.

또한 의사역행렬을 이용하여 수식(8)과 같이 구할 수 있다.

$$[W_m | B_m] = D \cdot [V^T | U] ([V^T | U]^T \cdot [V^T | U])^{-1} \quad (8)$$

2.2.2 보간 방법

보간은 먼저 저해상도 국소 부분을 codebook에서 가장 일치하는 codeword를 찾아 cluster, C_m 을 찾는다. 그리고 찾은 C_m 에서 W_m 과 B_m 을 이용하여 실제 Y_r 값을 찾아 확대될 영상에 대치하면 보간은 이루어 진다. LAM의 복원과정은 수식(9)과 같다.

$$\hat{X}_h = unV(W_m \cdot V(X_{n,r}) + B_m); X_{n,r} \in C_m \quad (9)$$

여기서 $unV(\cdot)$ 은 벡터화의 역연산이다.

그림 1에서 알 수 있듯이 각각의 입력에 대하여 고해상도 영상에서 같은 영역에 반복해서 출력을 나타내는 경우가 생긴다. 그러나 보간이란 고해상도의 실제 값을 추정하는 과정이기에 단지 평균값을 취한다.

2.3 결과 및 비교

학습영상의 실험은 두 개의 얼굴영상으로 하였으며 영상의 확대는 2배로 한다. 성능을 비교할 lena영상은 얼굴 부분의 131x129영상을 먼저 1/2 다운샘플한 후 학습된 신경망시스템에 입력하여 성능을 비교한다.

선형보간과 삼차 스플라인 보간방법, 신경방법에 관하여 수식 (10)과 같이 Mean-Square SNR측면에서 성능을 비교하겠다.

$$SNR = -10 \log_{10} \frac{E^2}{\max\{X(i,j)\}^2} \quad (10)$$

여기서 $E^2 = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M [\hat{X}_h(i,j) - X_h(i,j)]^2$ 이다

각각의 보간 방법에 대한 성능을 비교하였으며 그림3에서는 lena 영상에 대한 결과를 보였다.

복사 보간법 : SNR = 23.9504

선형 보간법 : SNR = 27.2172

삼차-convolution 보간법[5] : SNR = 27.3881

삼차-스플라인 보간법, $a=-1$: SNR = 26.9667

삼차-스플라인 보간법, $a = -0.578125$ (MAX) :

SNR = 27.362

삼차-스플라인 보간법, $a = -0.609375$ (MSE) : SNR = 27.347

신경망 보간법 : SNR = 36.1409

3. 결 론

우리는 기존의 스플라인 보간방법을 학습하였으며 하이퍼볼릭 탄젠트 모서리 모델에 대하여 최적인 커널을 찾았다. 기존의 연구결과와 비슷한 성능을 보였다

신경망을 이용한 영상 보간 방법은 미리 학습영상으로서 신경망을 훈련시킨 다음 보간할려는 영상을 신경망에 적용하여 보간하였다. 학습영상과 비슷한 특성의 영상은 잘 보간하였다. 따라서 특정 분야의 영상으로 학습하고 그 분야의 영상을 보간할 때 효과적으로 쓰일 수 있다. 그리고 특정분야의 영상의 압축에도 활용할 수 있을 것이다. 그러나 신경망을 이용한 방법은 학습 영상에 따라 성능의 차이가 나타나며 많은 연산을 필요하고 학습영상과 국소 부분에서의 특성이 상이한 영상을 복원하면 좋은 효과를 나타내지는 못한다.

(참 고 문 헌)

- [1] R.D.Dony and S.Haykin, "Neural Network Approaches to Image Compression", Proc. of IEEE, 83(2), Feb., pp288-303, 1995
- [2] H.S.Hou and H.C.Andrews, "Cubic splines for image interpolation and digital filtering", IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-26 pp.508-517, 1978
- [3] J.A.Parker and R.V.Kenyon and D.E.Troxel, "Comparison of Interpolating Methods for Image Resampling", IEEE Trans. medical imaging, vol.MI-2, NO.1, Mar, pp31-39, 1983
- [4] F.M.Candocia and J.C.Principe, "A Neural Implementaion of Interpolation with a Family of Kernels", ICNN'97, vol3, pp1506-1511, 1997
- [5] R.G.Keys, "Cubic Convolution Interpolation for Digital Image Processing", IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-29 pp.1153-1160, 1981

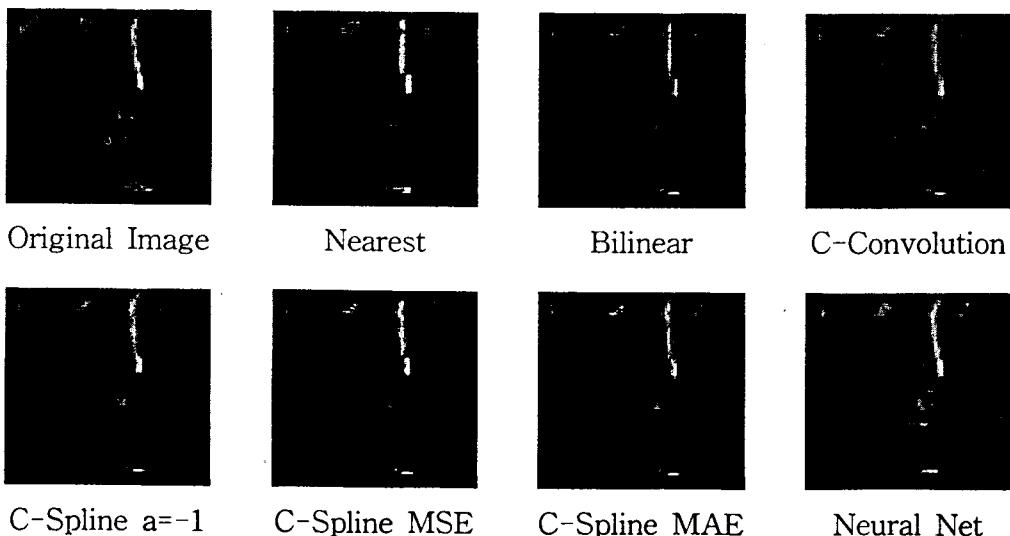


그림3 보간 방법에 대한 비교