

Pyramid 계층적 구조를 이용한 영상 합성 기법

정 석 윤 , 이 상 욱

서울대학교 공과대학 제어계측공학과

An Image Synthesis Technique based on the Pyramid Structure

Seok Yun Jung , Sang Uk Lee

Dept. of Control and Instrumentation Eng., Seoul National Uni.

ABSTRACT

We describe a texture image synthesis technique based on the AR (auto-regressive) model in this paper. Since the AR model is appropriate to a random texture field, we employ the pyramidal processing technique to decompose the texture containing some non-random local structure or non-homogeneity into random texture field.

Two examples are provided to demonstrate usefulness of the proposed method. Compare to the other synthesis techniques, the obtained images were more close to the original image subjectively.

1. 서 론

영상 부호화는 디지털화된 영상 정보에서 각 화소들간의 또는, 이웃하는 프레임간의 상관성(correlation)을 이용하여 영상 정보의 redundancy를 최소화하여, 방대한 양의 영상 정보를 최소로 줄이되, 원래의 영상을 충실히 재생할수 있도록하는데 그 목적이 있다. 영상 부호화는, 정보이론에 근거를 둔 제 1 세대 부호화 기법의 PCM(pulse code modulation), DPCM (differential PCM)등의 방법들과, 인간의 시각적 특성까지를 고려하는 제 2 세대 부호화 기법으로 나눌 수 있다. 제 2 세대 부호화 기법중에, contour와 texture를 이용한 부호화 기법이 있는데, 이것은, 영상 신호를 비슷한 성질(밝기, 색, 등)을 갖는 영역인 texture영역과 성질이 급격히 변하는 contour영역으로 분할하여 부호화하는 것이다. Texture[1,3]란, 영상에서 한 영역이 시각적으로 같은 영상내의 어느 다른 영역들과 구분하기 어려울때, 이 영역들을 하나의 texture영역이라 한다. 예로서, 잔디밭 이나 모래밭, 벽돌담, 나무나 가죽의 표면등이 대표적인 texture영상이다. texture영상은 다른 영상들과는 달리 영상의 각 화소값들을 모두 전송하거나, 각 화소값들을 직접 사용해서 영상 부호화해서 전송하기보다는, texture영상의 어떤 특성들을 알아내어 이 특성값들을 전송해서, 수신단에서 합성(synthesis) 영상을 만들어내는편이 전송해야할 정보량이 훨씬 작게 된다. 이때 사용되는 texture영상의 특성값들이란, texture영상을 어떤 modeling기법에 의해서 modeling했을때의

modeling parameter를 말한다. 이러한 texture영상은 다시 두 부류로 나뉘는데, 하나는 벽돌담이나, 밭집같이 같은 패턴이 규칙적으로 반복되는 영상들과, 또 하나는 가죽 표면이나, 잔디밭처럼 규칙적인 패턴은 찾아볼수 없지만, 화소값들의 통계적인 특성이 texture영역의 전체에 걸쳐서 거의 일정한 영상들로 구분할 수 있다. 전자의 경우는, 기본 패턴의 모양과 그것이 반복되는 규칙을 서술함으로써 texture영상을 표현할 수 있으므로 그 서술된 내용으로부터 처음의 texture를 다시 재생시킬수 있다, 이러한 경우를 구조적인 모델링[1](structural modeling) 이라한다. 후자의 경우는, 영상의 반복적인 규칙이 없고, 단지 통계적인 특성만을 가지고 texture를 표현해야 한다. 이러한 경우를 통계적 모델링[1](statistical modeling)이라한다. 본 논문에서는, texture영상의 통계적 모델링 방법으로서, 2 차원의 AR (auto-regressive) 모델링 방법을 이용하여, texture 영상을 AR process로 보았을 때의 AR 모델 계수로 나타내도록 하는데 중점을 둔다. 또한, texture 영상의 모델링과 그것에 따른 합성 영상의 화질을 개선하기 위해 Pyramid 구조를 도입 하였다.

2. texture영상의 통계적 성질을 이용한 modeling

Texture영상을 통계적 방법외에서 modeling하는 기법중에는, texture 영상을 하나의 stochastic process로서 modeling하여, 그 영상에 해당되는 모델링 계수를 구하는 방법이 있다. texture 모델링에 적용되는 stochastic process로는 주로 spatial AR 모델이 이용된다. 1차원 신호인 음성 신호에 대해서는 오래전부터 선형 예측 부호화라하여, 음성 신호 자체를 AR 모델로 모델링 하여 저전송률의 부호화에 적용해왔다. 음성 신호는 음성의 발생 과정에 대한 물리적인 모델링 결과로부터 AR 모델로의 유도가 가능하기때문에, 음성에 대한 AR 모델링 기법은 적절한 기법이라고 알려져있다. 이러한 AR 모델 기법을 영상 신호에도 적용하려면, 먼저 영상 신호들에 대한 해석이 이루어져야하는데, 영상 신호에 대한 해석은 현재로서는 적절한 해석 방법이 알려지지 못한 실정이다. 다만, 영상 신호는 주변 화소들간의 상관 관계가 매우 높다는 사실에 근거해서, 영상 신호에 대하여 AR 모델을 적용할 수 있게 된 것이다. 통계적 성질에 의한 texture 영상의 모델링 기법으로서, 또 다른

것은 주어진 texture 영상의 어떤 통계적 특성값 (texture feature)들을 구해서, 이 값들로부터 다시 texture 영상을 합성해내는 것들이 있다. 여기에 사용되는 통계적 특성값으로 무엇을 사용할 것인가에 대한 제안들이 여러가지 있다. D.D. Garber는 texture feature로서, 주어진 원래의 texture 영상의 일부(제안자의 실험에서는 원 영상의 1/4 에 해당하는 영역)를 전송하여, 수신단에서는 이 정보를 이용하여, 나머지 부분 (즉, 3/4 에 해당하는 영역)을 채워 넣는 방법을 제안하여, Best-Fit 모델[1]이라 하였다. 이 방법은 합성 영상의 화질 (즉, 원 영상과의 유사도)면에 있어서는 매우 뛰어난 성능을 보이지만, 계산량이 너무 많아 그리 효율적이지 못하다. A. Gagalowicz 등이 제안한 Sequential Synthesis 방법[2]은 주어진 texture 영상으로부터 0차, 1차, 2차 까지의 spatial statistics를 계산하여, 이 값들을 전송하고, 수신단에서는 전송되어온 0차, 1차, 2차의 spatial statistics와 거의 같은 값을 가지는 영상을 만들어내는 방법으로, 비교적 짧은 계산시간에 비해서 좋은 화질을 얻을 수 있으나, 화소의 밝기등급의 분해능(resolution)이 나쁘고, 거의 random noise에 가까운 영상들에만 잘 적용된다. 본 연구에서는, 이상의 통계적 모델중, 적은 양의 모델 계수로 texture를 나타낼 수 있는 AR 모델 기법을 개선하여, 합성 영상의 화질을 좋게 하는 방법에 대해 고찰하였다.

3. Non-causal AR 모델의 계수 추정[9]

Texture 영상에 대한 여러가지의 모델링 기법중 AR 모델링 기법은, texture 영상을 백색 가우시안 잡음 (white Gaussian noise)을 입력으로 하는 AR process로 간주하여, AR 모델링 계수로서, 주어진 texture 영상을 표현한다. 먼저, 주어진 texture 영상의 크기가 $M \times M$ 이고, 각 화소의 밝기값들이 $y(i,j)$ 로 표현될때, 편의상 (i,j) 를 2차원의 vector s 로 나타내고, r 또한 2차원의 vector 일때, AR 모델링 식은 식 (1) 과 같다.

$$y(s) = \sum_{r \in N} a(r) \times y(r+s) + g \times w(s), \quad s \in \Omega \quad (1)$$

여기서, N 은 neighbor set, Ω 는 $M \times M$ 크기의 영상의 영역을 나타내며, $w(.)$ 는 unit variance 의 white Gaussian noise 이다. 이때의 $a(.)$ 들이 바로 AR 모델의 계수들이다. 한편, neighbor set N 에 의해서 모델이 causal인가 noncausal인가가 결정되는데, 본 논문에서는 toroidal lattice 형태의 neighbor set를 취하는 noncausal 모델을 채용했다. AR 모델링 식을 다시 고쳐 쓴다면, 다음의 식 (2) 처럼 표현할 수도 있다.

$$g \times w(s) = y(s) - \sum_{r \in N} a(r) \times y(r+s), \quad s \in \Omega \quad (2)$$

이 경우는 $a(.)$ 와 $y(.)$ 가 모두 주어졌을때, $w(.)$ 를 구하는 식으로, 이때의 $w(.)$ 를 일반적으로, residual signal 또는, prediction error 라 한다. 이와는 반대로, $a(.)$ 와 $w(.)$, g 가 주어졌을 때는, 앞서의 과정과는 역으로, $y(.)$ 를 구할 수 있는데, 이것을 영상의 합성과정이라 한다. 즉, 주어진

영상 $y(.)$ 로부터 $a(.)$, g 를 구하는 과정이 모델링 (또는 analysis) 과정이고, $a(.)$, g 로부터 $w(.)$ 는 unit variance 의 random noise) $y(.)$ 를 구하는 과정이 합성과정이다. 이제, R.Chellappa가 제안한 AR 모델링 기법에 의해 모델링하고, 합성하는 과정을 설명하겠다.

3.1 AR 모델에 의한 영상 합성

주어진 영상 $y(.)$ 의 크기가 무한대가 아니고, 유한한 크기 ($M \times M$)이므로, AR 모델링에 의해 영상을 합성할때나, residual signal을 구하는 analysis 과정은, 모두 초기값 (또는 boundary value)을 필요로 하는데, 흔히 생각할 수 있는 경우에는 이 초기값을 0 (zero)으로 할 수도 있다. 그러나, texture 영상의 특성상 $M \times M$ 크기의 영상을 가로,세로 방향으로 규칙적으로 반복 배치시켜도, 영상의 가장자리에서의 통계적 특성값들이 불연속적으로 변하는 경우는 없다고 볼 수 있으므로, AR 모델링 적용시 별 문제가 되지 않는다. 따라서, 이와같이 texture 영상을 반복적으로 배열해서 AR 모델에 의해 영상을 analysis 또는 합성시의 초기 조건은 주어진 영상을 이용하게 된다. 이때의 AR 모델링 식은 다음의 식 (3) 처럼 된다.

$$y(s) = \sum_{r \in N} a(r) \times y(r \oplus s) + g \times w(s), \quad s \in \Omega \quad (3)$$

이때, \oplus 연산은 modulo M 을 뜻한다.

이 경우, 모델링 식을 residual signal에 대한 식으로 고쳐 쓰면,

$$y(s) - \sum_{r \in N} a(r) \times y(r \oplus s) = g \times w(s), \quad s \in \Omega \quad (4)$$

와 같이 되는데 이식은 다시 matrix 를 이용해서 표현하면 식 (5) 와 같이 된다.

$$B(a) y = g w$$

여기서, $B(a)$ 는 $a(.)$ 로 이루어진 $M^2 \times M^2$ 크기의 matrix, y , w 는 $y(.)$, $w(.)$ 를 각각 lexicographic 순으로 재배열한 $M^2 \times 1$ 크기의 matrix 이다.

이때의 $B(a)$ 는 block circulant matrix 로써 Fourier vector로 이루어진 matrix F 와 F^{-1} 에 의해서 대각화(diagonalize)시킬 수 있다[12]. 즉,

$$B(a) = F^{-1} D(a) F \quad (5)$$

여기서,

$$F = [((f_{kl}), \quad l=0, M-1), \quad k=0, M-1]$$

($M^2 \times M^2$ 의 matrix)

$$f_{kl} = [((t_{ij,kl}), \quad j=0, M-1), \quad i=0, M-1]^T$$

($M^2 \times 1$ 의 matrix)

$$t_{ij,kl} = \exp [-\sqrt{-1} \times 2\pi (ik+jl) / M]$$

또한, Fourier 변환의 성질에 의해

$$F^{-1} = F^* / M^2 \text{ 와 같이 표시할 수 있다.}$$

그리고, 대각 행렬 $D(a)$ 는 $B(a)$ 의 고유값 (eigenvalue) 으로 이루어진 행렬이다. 이러한 M^2 개의 고유값 μ_s 을 다음의 식 (6) 처럼 표현된다.

$$\mu_s = (1 - 2 a^T \Phi_s) \quad (6.a)$$

$$\Phi_s = [\exp [\sqrt{-1} \times 2\pi s^T r] , r \in N]^T \quad (6.b)$$

이와같은 식들을 이용해서 다시 정리하면,

$$y = g F^{-1} D^{-1} F w \quad (7)$$

와 같이 된다.

한편, $F w$ 는 $M \times M$ point의 2 차원 DFT를 뜻하므로, 결국 영상의 합성 과정은 $M \times M$ point의 2 차원 DFT를 이용하게 된다.

3.2 AR 모델링의 계수 추정

본 논문에서는 AR 모델의 계수 추정 방법으로, R.Chellappa가 제안한, toroidal neighbor 에 대한 ML (maximum likelihood) estimation 방법을 사용했다. 즉, 모델링 식에서, $w(\cdot)$ 가 가우시안이므로 $y(\cdot)$ 의 pdf는

$$p(y|a, g) = [|B(a)| / (2\pi g^2)^{M^2/2}] \times \exp [-(2g^2)^{-1} \sum \{ y(s-a^T z(s))^2 \}] \quad (8)$$

이 되고,

likelihood function은

$$\log(p(y|a, g)) = \log|B(a)| - (M^2/2) \log(2\pi g^2) - (2g^2)^{-1} \sum \{ y(s-a^T z(s))^2 \} \quad (9)$$

여기서,

$$z(s) = [y(r \odot s), r \in N]^T$$

이러한 likelihood function을 최대화시키는 $a(\cdot)$, g 를 구하기 위해 nonlinear programming기법을 사용해서 최종적으로 AR 모델의 계수를 구한다. 그리고, 입력 영상에 적합한 모델 차수를 결정하는 방법으로는, 합성 영상과 원 영상과의 시각적 유사도를 기준으로 해서 가장 비슷한 합성 영상을 얻을수있는 경우를 그 영상에 적합한 모델차수로 결정했다.

4. Pyramid 구조를 이용한 영상 합성[10]

Texture 영상에 AR 모델을 적용하기 위해서는 우선, texture 영상이 stationary process 이어야 하고, 또한 적은 양의 계수로 모델링 하려면 영상의 화소간의 상관도가 거리가 멀어지면 급격히 떨어지는 즉, fine structure의 texture 영상이어야 한다. 후자의 조건은 입력 영상을 적절히 선택하면 잘 만족시킬 수 있으나, 전자의 조건은, 영상 취득시의 조명의 불균일 등으로 인해서, texture영상이 균질(homogeneous) 한 영상이라 할수 없게 된다. 따라서, 입력 영상이 불균일 하다면, 모델링을 하기에 앞서 영상을 균질화시킬 필요가 있다. 예로서, D.D. Garber는 이런 경우에 statistical differencing 을 행해서, 불균질의 영상을 모델링에 적합한 영상으로 바꾸어 주었다 [1]. 그런데, 불균일한 영상을 high pass signal 만을 취해보면 거의 균질의 신호로 볼수 있게 된다. 즉, texture영상은, high pass signal 을 취하면 원 영상에 비해 균질의 영상으로 되

로 AR 모델링에 적합할 것이다. 그런데, Pyramid 구조의 각층의 신호는 모두, high pass signal이므로, AR 모델링의 요구조건을 충족시키는 신호로 나누어 주게 된다. 이런 관점에서, Pyramid 의 각층의 신호에 각각 AR 모델을 적용해서, 입력 영상이 불균질의 영상이거나, 또는 약간의 구조적 특성이 존재해도, 이 영상들을 분해 (decomposition) 해서 원 영상에 비해 상관도가 낮은 영상들로 만들어서 각각에 대해서 AR 모델을 적용하는 방법을 제안하게 되었다. 입력 영상으로부터 Pyramid 신호를 만드는 과정은 그림 (1) 에 나타내었다.

5. 실험 결과

실제 texture 영상에 대한 실험은 다음의 순서로 진행되었다. 먼저, 영상 취득장치로부터 실제의 texture영상을 얻고, 이 영상에 대해 Pyramid의 각층의 신호를 만들어 낸다. 이때의 영상의 크기는 계산상의 이유로, 64×64 크기의 영상을 입력 영상으로 하였다. 다음, Pyramid 각층의 영상은 각각, 64×64 , 32×32 , 16×16 의 것이 되는데, 64×64 , 32×32 의 두 신호에 대해서, AR 모델링을 적용하고, 16×16 의 신호는 정보의 양이 많지 않으므로, 이 층의 신호는 신호 자체를 그대로 전송한다. 따라서, 입력영상에 관하여, 전송해야할 정보의 양은, 16×16 크기의 Pyramid구조의 가장 작은 크기의 신호와 32×32 , 64×64 크기의 신호에 대한 AR 모델 계수들이다. AR 모델링의 차수는 고정된 것이 아니고, 몇가지의 다른 차수로 모델링 (본 논문에서는, 2, 4, 6, 10, 12, 14, 18, 22, 24 의 9 가지) 한후 이중 합성영상의 화질이 시각적으로 가장 뛰어난 것을 해당 영상의 모델 차수로 결정하였다. 실험의 예로서, Brodatz 의 사진들중 D11 로 표시된 섬유의 사진과, D84 로 표시된 raffia영상을 사용했다. 이 중, 섬유의 영상은, 영상내의 밝기가 고르지 못하고 (세로방향의 strip 문양이 존재), 빗금방향의 선이있는영상으로서 이것을 바로 AR모델로 나타내기에는 부적합한 경우이다. 이런 영상에 대해서, Pyramid 각층의 신호를 만들어내고, 각각의 신호에대해서, AR 모델을 적용하여, 최종적으로는 AR 모델계수와, Pyramid의 가장작은 크기의 신호를 이용해서, 합성영상을 만들어 내었다. 이것들을 그림 (2) 에서부터 그림 (7) 까지 보여줬는데, 그림 (2.a)는 원래의 섬유 영상이고, 그림 (2.b)는 이에대한 Pyramid구조의 각 층의 영상이다. 그리고, 그림 (2.c)는 합성 영상으로서, Pyramid각층에 대한 합성영상이고, 그림 (2.d)는 최종적으로 합성된 것이다. 또한, 본 논문의 방법과 비교하기위해서, 원 영상에대해 직접 AR 모델을 적용했을 때, AR 모델의 계수의 양(차수) 에 따른 합성 영상들을 그림 (3) 에 보였으며, 그림 (4) 는 합성 영상을 만들때의 excitation 신호로서 residual 신호를 1 bit 의 Max 양 차화한 것을 사용하여 합성한 것이다. 그림 (2) 에서부터, 그림 (4) 까지 보여주시 알수 있듯이, 본 논문에서 제안한 모델링 기법이 양자화한 residual신호를 사용한 기법보다도 더 나은 합성 영상을 얻었다. 또다른 예로서, raffia 영상에대한 실험 결과도, 이러한 결과를 보여주는데, 그림 (5) 에서부터, 그림 (7) 까지 보여졌다.

6. 결 론

본 논문에서 제시한, Pyramid 구조와 AR 모델을 결합한 형태의 texture 모델링 방법은, 입력 영상이 불균질(non-homogeneous)한 영상이라도, Pyramid 구조로서 이것을 균질한 영상들로 바꾸어준 후, 거기에 AR 모델을 적용함으로써, 보다 정확한 모델링을 기대할수 있도록 하였다. 만약, 입력 영상의 상태가 거의 random noise에 가까운 경우라면, Pyramid 구조가 필요 없이 바로 AR 모델을 적용할수도 있다. 그러나, 입력 영상이 불균질의 것이거나, 어느정도의 구조적 특성이 존재하는 경우라면, 제안한 바와 같이 Pyramid구조를 이용하는 편이 합성 영상의 화질면에서 우수하다.

7. 참고 문헌

- [1] D.D. Garber, "Computational Models for Texture Analysis and Texture Synthesis," USCIPI, 1981.
- [2] A. Gagalowicz, S.D. Ma, "Sequential Synthesis of Natural Textures," Computer Vision, Graphics, and Image Processing 30, pp289-315, 1985.
- [3] A. Rosenfeld, A.C. Kak, *Digital Image Processing*, Vol. 2, Academic Press, New York, 1976.
- [4] A.K. Jain, "Advances in Mathematical Models for Image Processing," Proc. of IEEE, Vol.69, No.5, May 1981.
- [5] S.J. Orfanidis, *Optimum Signal Processing*, McMillan Publishing Co., 1985.
- [6] A.A. Giordano, F.M. Hsu, *Least Square Estimation with Application to Digital Signal Processing*, a Wiley Interscience Publication, John Wiley and Sons, 1985.
- [7] S. Ranganath, A.K. Jain, "Two-Dimensional Linear Prediction Models - Part I : Spectral Factorization and Realization," IEEE ASSP Vol.33, No.1, Feb. 1985.
- [8] T.L. Marzetta, "Two-D Lenear Prediction Autocorrelation Arrays, Minimum - Phase Prediction Error Filters, and Reflection Coefficient Arrays," IEEE ASSP Vol.28, No.6, Dec. 1980.
- [9] R. Chellappa, "Stochastic models in image analysis and processing," Ph.D. dissertation, Purdue University, Lafayette, IN, Aug. 1981.
- [10] 박 섭형, 이 상욱, "영상 피라미드를 위한 최적 Half-Band FIR 필터," 전자공학회 논문지, 제 25 권 제 7 호, 1988년 7 월.
- [11] D.H. Ballard, C.M. Brown, *Computer Vision*, Prectice-Hall, 1982.

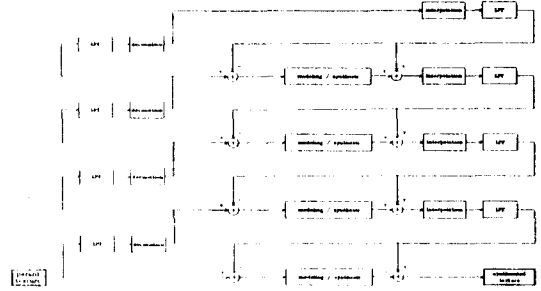


그림 1. 제안한 방법의 블록도

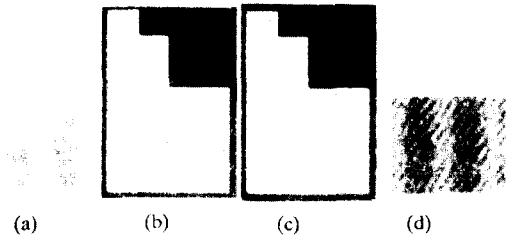


그림 2. Cloth 영상에 대한 실험 결과

- (a) 원래의 영상 (64 × 64)
- (b) pyramid 각 계층의 영상
- (c) pyramid 각 계층에 대한 합성 영상
- (d) 최종 합성 영상 (64 × 64)

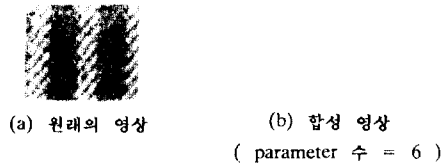


그림 3. Cloth 영상에 직접 AR 모델을 적용했을 때

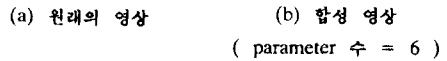


그림 4. Cloth 영상에 대해, 1 bit 양자화된 residual 사용

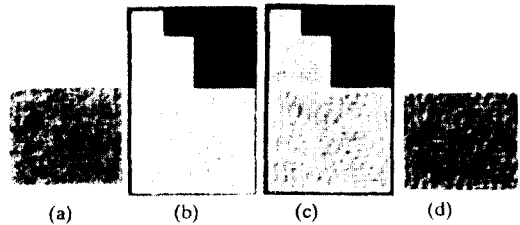


그림 5. Raffia 영상에 대한 실험 결과

- (a) 원래의 영상 (64 × 64)
- (b) pyramid 각 계층의 영상
- (c) pyramid 각 계층에 대한 합성 영상
- (d) 최종 합성 영상 (64 × 64)

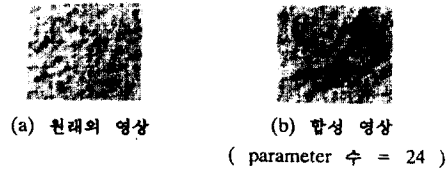


그림 6. Raffia 영상에 직접 AR 모델을 적용했을때

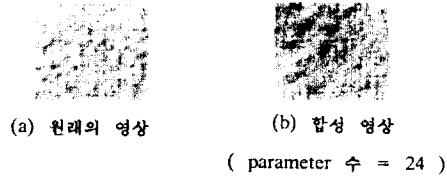


그림 7. Raffia 영상에 대해, 1 bit 양자화된 residual 사용