

동영상에서의 확장된 시공간 적응적 Auto-regressive 모델의 연구

(Extended Adaptive Spatio-Temporal Auto-Regressive Model for Video Sequence)

杜錫柱*, 姜文基**

(Seok-Joo Doo and Moon-Gi Kang)

요약

본 논문에서는 보다 정확한 선형예측을 위해 시공간적응적 기반영역에 바탕을 둔 확장된 auto-regressive(AR) 모델을 제안한다. 기존의 AR 모델에서 영상 전체에 걸쳐 사용되는 직사각형 형태의 기반영역은 영상의 경계선 영역에서는 더 이상 정상상태(stationarity)의 조건을 만족시키지 않음으로써 경계선 영역에서 예측오차가 증가하는 결점을 갖는다. 그래서 제안된 방법은 AR 모델에서 시공간적응적 기반영역이라 불리는 정상상태의 기반영역을 구성하는데 중점을 둔다. 시공간적응적 기반영역은 영상의 경계선 특성과 관련되는 공간적응적 기반영역과 시간축의 불연속 개념과 관련되는 시간적응적 기반영역으로 구성되어진다. 제안된 AR 모델은 동영상 복원 실험간 좀 더 정확한 모델 파라미터를 산출하였을 뿐만 아니라 복잡한 계산과정을 단순화하는 이점을 가진다.

Abstract

In this paper, a generalized auto-regressive(AR) model is proposed for linear prediction based on adaptive spatio-temporal support region(ASTSR). The conventional AR model suffers from the drawback that the prediction error increases in the edge region because the rectangular support region of the edge does not satisfy the stationary assumption. Thus, the proposed approach puts an emphasis on the formulation of a spatio-temporally adaptive support region for the AR model, called ASTSR. The ASTSR consists of two parts: the adaptive spatial support region(ASSR) connected with edges and the adaptive temporal support region(ATSR) related to temporal discontinuities. The AR model based on ASTSR not only produces more accurate model parameters but also reduces the computational complexity in the motion picture restoration.

I. 개요

AR 모델의 중요한 응용분야는 정지영상 또는 동영상

(motion picture)에서 손실된 데이터를 판별하고 보간하는 영상의 복원이다. 영상 복원을 위한 AR 모델의 잡음 판별 및 보간법에서는 먼저 잡음 판별 프로세스에 의해 잡음의 위치를 판별하고, 판별된 데이터에 대해서만 적절한 데이터 값으로 보간하는 방법을 취하게 되며 이러한 방법은 여러 형태로 연구되어 왔다^[13].

이러한 기존의 AR 모델에서는 예측을 위한 기반영역(support region)으로 직사각형 형태의 영역을 선택하여 편의상 영상 전체에 대하여 적용을 하였는데 이 경우 영상의 경계선 영역에서 예측오차가 상대적으로 증

* 正會員, 陸軍第3士官學校 電子工學科

(Dept. of Electronic Eng., Korea Third Military Academy)

** 正會員, 延世大學校 電子工學科

(Dept. of Electronic Eng., Yonsei University)

接受日字:1999年1月22日, 수정완료일:1999年9月3日

가하는 결점을 갖는다. 그 이유는 경계선 영역에서는 직사각형 형태의 기반영역이 더 이상 정상상태(stationarity)를 만족시키지 못한다는 데에 있다. 일반적으로 정상 프로세스(stationary process)는 WSS(Wide-Sense Stationary)를 의미하므로 만일 자기상관계수 값이 절대적 시간이나 공간에 의존하지 않는다면 AR 모델에서의 정상상태 문제는 단순화될 수 있다. 그러나 일반적으로 주어진 영상이 언제나 정상상태를 만족시키지는 않는다. 특히 경계선 영역은 거의 비정상상태(nonstationarity)를 지니게 되며, 정상상태를 요구하는 AR 모델에서 직사각형 형태의 기반영역을 경계선 영역에 그대로 적용하는 것은 어렵다. 이런 비정상상태의 문제해결을 위해 quadtree 분할방법을 이용한 연구^[4]와, 비정상상태의 특성에 따라 AR 모델 계수들에 적절한 가중치를 부여하는 방법이 연구되어왔다^[5]. 본 논문은 비정상상태의 문제 해결을 위해 시공간적응적 기반영역에 바탕을 둔 확장된 AR 모델을 제안한다. 이 모델은 비정상상태 영역에서도 예측필터에 필요한 최적화된 모델 파라미터들을 산출한다.

논문의 구성은 다음과 같다. 2장에서는 시퀀스에서의 AR 모델에 대해 설명하고, 3장에서는 시공간적응적 기반영역의 설정에 대해 기술한다. 4장에서는 제안된 AR 모델을 통한 동영상 복원의 응용 및 실험 결과를 제시하며, 마지막 장에서는 결론을 제시한다.

II. 시공간 Auto-Regressive 모델

전통적인 선형예측방법은 사전 설정된 기반영역에 속하는 화소값들에 적절한 가중치가 부여된 선형 조합에 바탕을 두고, 주어진 영상의 한 화소값을 예측하는 것을 의미한다^[1,6]. 이러한 3차원(3-D) AR 모델은 다음과 같이 표현될 수 있다.

$$I(x, y, m) = \sum_i \sum_j \sum_n a(i, j, n) I(x+i+dx, y+j+dy, m+n) + e(x, y, m), (i, j, n) \in \hat{V}. \quad (1)$$

주어진 식에서 $I(x, y, m)$ 은 m 번째 프레임의 위치 (x, y) 에서의 화소값을 나타내며, $a(i, j, n)$ 은 모델 파라미터이다. 또한 $[dx, dy]$ 는 프레임 과 사이의 움직임 벡터를 의미하며 $e(x, y, m)$ 은 예측오차이다. \hat{V} 은 시공간 기반영역으로서 일반적으로 다음과 같이 나타내어진다.

$$\hat{V} = \{ -p \leq i \leq p, -q \leq j \leq q, -r \leq n \leq r, (i, j, n) \neq (0, 0, 0) \} \quad (2)$$

식 (1)로부터, 예측값 $\hat{I}(x, y, m)$ 은

$$\hat{I}(x, y, m) = \sum_i \sum_j \sum_n a(i, j, n) I(x+i+dx, y+j+dy, m+n) + (i, j, n) \in \hat{V} \quad (3)$$

이다. 식 (3)에 포함되어 있는 AR 모델 파라미터를 산출하기 위하여 식 (1)과 (3)로부터 다음과 같이 주어지는 예측오차를 고려해볼 수 있다.

$$e(x, y, m) = I(x, y, m) - \hat{I}(x, y, m) \quad (4)$$

구하고자 하는 모델 파라미터는 식 (4)에서 예측오차의 제곱값, 즉 $e^2(x, y, m)$ 이 최소화되는 상태에서의 파라미터들로 간주될 수 있다.

$$\frac{\partial E[e^2(x, y, m)]}{\partial a(i, j, n)} = 0 \quad (5)$$

직교성(orthogonality principle)을 적용하여 식 (5)를 전개하면 모델 파라미터를 구하기 위한 다음과 같은 정규 방정식(normal equation)을 얻게 된다.

$$R\mathbf{a} = \mathbf{b} \quad (6)$$

여기서 R 은 Toeplitz행렬을 이루며, \mathbf{a} 는 AR 모델 파라미터로 이루어진 벡터를, \mathbf{b} 는 자기 상관 계수들로 이루어진 벡터를 각각 나타낸다.

III. 시공간적응적 기반영역의 구성

시공간적응적 기반영역은 공간적응적 기반영역과 시간적응적 기반영역으로 구분된다.

1. 공간적응적 기반영역

공간적응적 기반영역은 사전에 주어진 기반영역으로부터 공간적으로 정상적인 기반영역의 구성에 유효한 화소들을 경계선의 특성과 관련하여 선택함으로써 구성된다. 여기서 주어진 기반영역이란 기존 AR 모델에서 사용되는 임의의 직사각형 형태의 기반영역을 의미한다. 공간적응적 기반영역을 구성하는 판별기는 다음과 같다.

$$D = \begin{cases} 1, & \text{if } PD(k, l) \leq \tau_s, \quad (k, l) \in \hat{W} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

$$PD(k, l) = |u(x, y) - u(k, l)| \quad (8)$$

식에서 $u(x, y)$ 는 모델에 의해 예측되는 값이며, $u(k, l)$ 은 사전에 설정된 2차원 기반영역 \hat{W} 내 위치 (k, l) 에서의 화소값을 나타낸다. 판별기에 의해 기반영역 내의 어느 한 화소는 해당 화소값과 예측되어질 화소값의 차가 임의의 임계값 τ_s 보다 작을 때 새로운 기반영역의 구성에 유효한 화소로 간주된다. 그림 1에서 판별기는 사전에 설정된 3×3 기반영역으로부터 7개의 화소들을 정상적인 공간적응적 기반영역 \hat{W}_s 으로 구성하고 있다.

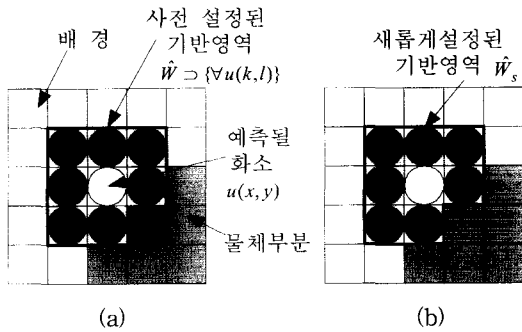


그림 1. (a) 고정된 기반영역
(b) 공간적응적 기반영역
Fig. 1. (a) fixed support region
(b) adaptive spatial support region.

2. 시간적응적 기반영역

시간축 신호의 비정상상태를 해결하기 위해서는 일반적으로 시퀀스의 움직임을 보상함으로써 비정상적 요소를 제거한다. 그러나 부가적인 움직임 판별기 (motion estimator)의 사용은 많은 계산량을 요구하므로 본 논문에서는 움직임 벡터를 사용할 필요가 있다고 판정되는 영역에 대해서만 부분적으로 움직임 벡터를 구하는 방법을 취한다.

시간적응적 기반영역은 움직임의 고려여부에 따라 엄밀한 의미의 시간적응적 기반영역과 움직임이 고려된 시간적응적 기반영역의 두 가지 형태로 분류된다. 이 장에서는 먼저 움직임의 존재를 판별하는 방법에 대해 설명하고, 두 가지 형태의 시간적응적 기반영역에 대하여 설명한다.

1) 움직임변화의 판별
움직임변화 판별(change detection)은 움직임으로 인해 주의가 요구되는 부분을 결정함으로써 움직임 판별기를 사용해야 되는 부분을 결정하는데 목적을 둔다.

움직임변화 판별을 위해 두 프레임의 같은 위치에서의 일정 영역에 해당하는 화소값들의 통계적 특성에 기반을 둔 비교 방법을 사용할 수 있다. M 을 $m \times n$ 의 직사각형 크기를 갖는 임의의 영역이라 하고 μ 와 σ 를 각각 M 에 해당하는 화소값들의 평균과 분산이라 할 때, 움직임 비교 파라미터 λ 는 다음과 같이 정의된다^[7].

$$\lambda = \frac{\left[\frac{\sigma_i + \sigma_j}{2} + \left(\frac{\mu_i - \mu_j}{2} \right)^2 \right]}{\sigma_i \times \sigma_j} \quad (9)$$

여기서 기호 밑의 첨자는 각 프레임을 나타낸다. 만일 두 프레임 i 와 j 사이의 설정된 영역에 움직임이 존재하지 않는다면 비교 파라미터는 최소값 1을 가지며, 움직임의 크기가 커질수록 비교 파라미터의 값은 증가하는 특성을 갖는다. 이런 특성을 이용하여 임의의 임계값 τ_i 에 따른 움직임변화 판별기는 다음과 같다.

$$FD_{ij}(x, y) = \begin{cases} 1, & \text{if } \lambda > \tau_i \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

2) 엄밀한 의미의 시간적응적 기반영역과 움직임을 고려한 시간적응적 기반영역
엄밀한 의미의 시간적응적 기반영역은 공간 영역에서의 경계선 개념을 시간 영역에서의 시간적 불연속 개념으로 그대로 확장함으로써 얻어진다. 그래서 시간축 정보는 시간적 불연속이 존재하지 않는 경우에 사용 가능하다. 엄밀한 의미의 시간적응적 기반영역을 구성하는 과정은 다음과 같다.

- 1단계 : 현재의 프레임에서 공간적응적 기반영역 \hat{W}_s 를 설정한다.
- 2단계 : 기존의 기반영역 \hat{W} 에 대해 양방향으로 움직임 변화 판별을 적용하고, 작은 비교 파라미터를 갖는 프레임을 기반 프레임으로 선정한다.
- 3단계 : 만일 움직임이 존재하지 않는다면 선정된 기반 프레임내에서의 시간적응적 기반영역 \hat{W}_t 는 공간적응적 기반영역 \hat{W}_s 와 같은 크기 및 형태로 설정된다.

그러나 만일 움직임이 존재한다면 시간적응적 기반영역은 정의되지 않는다.

그러나 움직임을 고려한 시간적응적 기반영역은 언제나 시간적응적 기반영역을 구성하며, 따라서 AR 모델을 이용한 선형예측에 시간축 정보를 항상 적용할 수 있다는 장점을 가진다. 즉, 움직임을 고려한 시간적응적 기반영역을 구성하기 위해서, 움직임변화 판별의 결과에 따라 만일 두 프레임간에 움직임이 존재하지 않는다면 시간적응적 기반영역은 기반 프레임 내에서 공간적응적 기반영역과 같은 위치에 같은 형태로 설정되게 된다. 그러나 만일 움직임이 존재한다고 판별되면 해당 위치에 대해 움직임 판별기를 적용하여 움직임 벡터를 구한 후, 이 움직임 벡터를 적용하여 기반영역을 구성하게 된다. 그림 2는 움직임이 없는 상태에서 구성될 수 있는 일반적인 시간적응적 기반영역의 구성예를 보여준다. 공간적응적 기반영역 \hat{W}_s 과 시간적응적 기반영역 \hat{W}_t 의 합은 시공간적응적 기반영역 \hat{V}_{st} 을 이룬다.

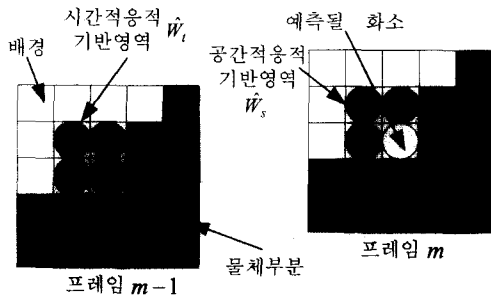


그림 2. 움직임이 없는 상태하에서의 시간적응적기반 영역의 구성

Fig. 2. Construction of adaptive temporal support region without motion.

IV. 응용 : 동영상 복원

오래된 영화 필름들은 먼지나 긁힘 등 여러 가지 잡음에 의해서 열화되는 특성을 가지며, 이중 전형적인 잡음은 흔히 'dirt and sparkle'이라 불리는 'bright and dark flashes' 현상이다^[1,6]. 이러한 잡음의 제거를 위해 시공간적응적 기반영역에 바탕을 둔 3-D AR 모델을 적용하였다.

1. 복원 모델을 위한 시공간적응적 기반영역

동영상 복원에서는 움직임이 존재하는 경우에 대해서도 시간적응적 기반영역을 구성해야 하는 필요성을

가진다. 왜냐하면 현 프레임내의 공간적응적 기반영역 만으로는 손실된 화소값을 예측하는데 충분한 기반영역을 구성할 수 없는 경우가 존재할 뿐만 아니라 잡음에 의한 시간적 불연속은 손실된 데이터를 판별하는데 유용한 정보로 사용될 수 있기 때문이다. 그래서 본 응용에서는 움직임을 고려한 시공간적응적 기반영역을 바탕으로 하는 3-D AR 모델을 적용하였다.

그러나 잡음에 의해 손실된 데이터가 현재 프레임에 존재할 경우 움직임을 고려한 시공간적응적 기반영역을 적용하는 데에는 다음과 같은 몇 가지 고려사항이 존재한다.

i) 식 (9)에서 비교 파라미터 λ 는 큰 값을 갖게 된다. 그 결과 판별기는 움직임이 존재하지 않음에도 불구하고 해당 잡음을 움직임으로 간주하게 된다. 이 경우 앞 프레임과 뒤 프레임을 대상으로 하는 움직임변화 판별의 적용이 필요하다. 실제 움직임이 존재한다면 여기서 구한 비교 파라미터 역시 큰 값을 갖게 될 것이다.

ii) 현재 프레임 내에 구성된 공간적응적 기반영역은 경계선의 특성에 관련되기 보다는 잡음과 관련되게 된다. 이 경우 공간적응적 기반영역을 현재 프레임이 아닌 기반 프레임에서 먼저 설정해야 한다. 물론 잡음의 랜덤한 특성상 연속된 프레임의 동일 위치에 잡음이 발생하지는 않는다고 간주할 수 있다.

2. 손실된 데이터의 판별 및 복원

시공간적응적 기반영역 \hat{V}_{st} 에 바탕을 둔 3-D AR 모델의 방정식은 다음과 같이 표현된다.

$$I(x, y, m) = \sum_i \sum_j \sum_n a(i, j, n) I(x+i+dx, y+j+dy, m+n) + e(x, y, m), (i, j, n) \in \hat{V}_{st}. \quad (11)$$

식 (11)로부터 손실된 데이터의 위치는 임의의 임계값을 아래와 같이 위치 (x, y, m) 에서의 예측오차의 제곱값에 적용함으로써 가능하다.

$$e_d = (I(x, y, m) - \hat{I}(x, y, m))^2. \quad (12)$$

손실된 데이터의 위치를 판별하는 기본 원리는 원래의 영상 $I(x, y, m)$ 은 AR 모델을 따르지만 잡음으로 손실된 데이터는 그렇지 않다는 것이다.

손실된 데이터의 복원 과정은 다음과 같은 3단계로 구성될 수 있다^[3]. 첫째, 손실되었다고 판별된 화소값들

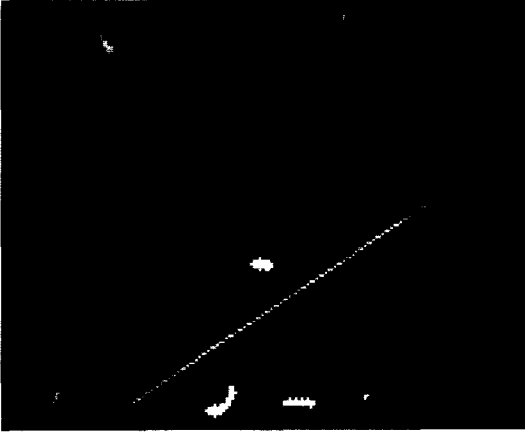


그림 3. 열화된 Salesman 영상(#3)
Fig. 3. Degraded frame 3 of Salesman.



그림 4. 복원된 Salesman 영상(#3)
Fig. 4. Restored frame 3 of Salesman.

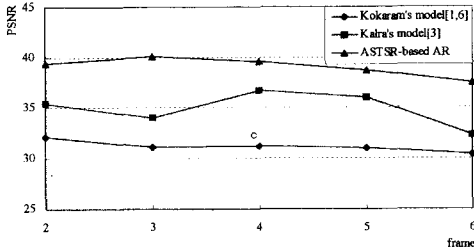


그림 5. 복원된 Salesman 영상의 PSNR 비교
Fig. 5. PSNR values of restored Salesman.

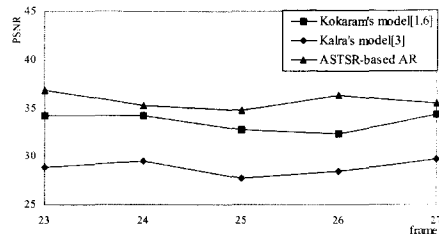


그림 6. 복원된 Hall Monitor 영상의 PSNR 비교
Fig. 6. PSNR values of restored Hall Monitor.

은 Gaussian weighting function(GWF)에 의해 가중치가 부여된다. 둘째, 식 (6)을 통하여 잡음에 의해 영향을 받지 않은 새로운 AR 모델 파라미터가 구하여진다. 마지막으로, 손실되었다고 판별된 데이터는 새로운 모델 파라미터에 의한 예측값 $\hat{I}(x, y, m)$ 으로 대체된다.

3. 실험 결과

열화된 영상 시퀀스 복원 실험을 위해 176×144 의 크기를 갖는 시퀀스 Salesman과 Hall Monitor에 대하여 [1]에서 제시된 잡음 발생 방법에 따라 임의의 크기와 형태를 갖는 잡음을 추가하였다. 그러나 잡음의 일반적인 특성을 고려하여 잡음의 크기는 255로 고정하였으며, 긁힘(scratch)의 효과적 표현을 위해 임의의 선을 추가하였다. 영상내 잡음 비율은 평균적으로 1.0%로 설정하였다. 제안된 모델의 성능은 기존의 Kokaram 9:00/0:9 AR 모델^[1]과 Kalra의 Gaussian weighting 3-D AR 모델^[3]로서 비교되었다. 실험간 필요한 움직임 판별기로는 블록 매칭 방법을 적용하였다.

그림 3과 4는 Sales 시퀀스의 3번째 프레임으로서 각각 열화된 영상과 제안된 모델을 통해 복원된 영상 ($\tau_s=2, \tau_t=e_d=10$)을 나타낸다. 표 1은 잡음을 판별한 결과를 그림 5와 6은 복원된 프레임들의 PSNR 비교를 나타낸다. 제안된 모델은 복원 프로세스내의 모델 파라미터를 구하는 과정에서 정상적 기반영역을 제공함으로써 복원 영상의 PSNR에서 타 모델들에 비해 성능 향상을 가져왔다. 더불어 기존의 9:0 AR 모델은 차수가 9인 반면, 실험간 제안된 모델의 평균 차수는 6.3으로서 그만큼 풀어야 되는 방정식의 수 면에서 이점을 가지며, 또한 움직임 판별 및 보상간에서 갖는 많은 계산량의 단점을 줄여 주었다.

V. 결 론

본 논문에서는 시공간적응적 기반영역에 바탕을 둔 확장된 AR 모델을 제안하였다. AR 모델에 시공간적응적 기반영역을 적용하는 목적은 영상의 경계선이나 시퀀스

표 1. 여러 AR모델에 의한 잡음판별결과
Table 1. Detection results with various AR models.

Detection Model	Salesman, #3		Hall Monitor, #26	
	%C	%F	%C	%F
Kokaram's 9:0:9 model[1] ($e_d=10$)	99.10	2.25	79.47	5.23
Kalra's 9:0:9 model[3] ($e_d=10$)	99.10	3.28	80.00	8.60
ASTSR-based AR ($\tau_s, \tau_r = s, e_d=10$)	99.55	1.61	90.53	4.52

%C=(no. of correctly detected pixels/total no. of missing pixels)*100,

%F=(no. of incorrectly detected pixels/total no. of pixels in a frame)*100.

의 시간적 불연속으로부터 정상적 특성을 갖는 기반영역을 보장하는데 있다. 시공간적응적 기반영역은 예측되는 화소와 높은 상관성을 갖는 화소들로 공간적응적 기반영역을 구성하고, 이 기반영역을 바탕으로 정상상태를 갖는 시간축으로 확장됨으로서 형성되어진다.

제안된 AR 모델은 사용된 기반영역이 정상적이라는 점에서 최적화되었다고 할 수 있으며, 따라서 동영상 복원 실험에서 기존의 모델에 비해 더 정확한 모델 파라미터를 산출할 뿐만 아니라 더불어 계산량을 효과적으로 줄이는 이점을 갖는다.

참 고 문 헌

[1] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald

and P. J. W. Rayner, Detection of Missing Data in Image Sequences, *IEEE Trans. Image Processing*, vol. 4, no. 11, pp.1496-1508, Nov. 1995.

[2] W. B. Goh, M. N. Chong, S. Kalra and D. Krishnan, Bi-directional 3-D Auto-Regressive Model Approach to Motion Picture Restoration, *IEEE ICASSP'96*, pp.2277-2280, May 1996.

[3] S. Kalra, M. N. Chong and D. Krishnan, A New Auto-Regressive Model-Based Algorithm for Motion Picture Restoration, *IEEE ICASSP '97*, vol.4, pp.2557-2560, Apr. 1997.

[4] P. Strobach, Quadtree-Structured Linear Prediction Models for Image Sequence Processing, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp.742-747, July 1989.

[5] S. N. Efstratiadis and A. K. Katsaggelos, Nonstationary AR Modeling and Constrained Recursive Estimation of the Displacement Field, *IEEE Trans. Circ. Syst. For Video Tech.*, vol. 2, pp.334-346, Dec, 1992.

[6] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald and P. J. W. Rayner, Interpolation of Missing Data in Image Sequences, *IEEE Trans. Image Processing*, vol. 4, no. 11, pp.1509-1519, Nov. 1995.

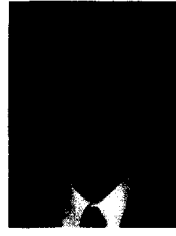
[7] R. Jain, R. Kasturi and B. G. Schunck, *Machine Vision*, McGraw-Hill, 1995.

저 자 소 개



杜錫柱(正會員)

1995년 3월 육군사관학교 전산학과 이학사. 1999년 2월 연세대학교 전자공학과 석사. 1999년 8월~현재 육군제3사관학교 전자공학과 강사. 주관심분야는 영상 복원, 디지털 비디오 처리등임.



姜文基(正會員)

1986년 2월 서울대학교 전자공학과 학사. 1988년 2월 서울대학교 전자공학과 석사. 1994년 2월 Northwestern University 졸업 공학 박사. 1994년 3월~1994년 10월 Northwestern University Research Fellow.

1994년 11월~1997년 2월 University of Minnesota, Duluth 조교수. 1997년 3월~1999년 2월 연세대학교 전자공학과 조교수. 1999년 3월~현재 연세대학교 전자공학과 부교수. 주관심분야 : 영상복원, 초고해상도 영상 복원, 디지털 비디오 처리