

다해상도 블럭 정합기법을 이용한 스테레오 동영상의 부호화

정희원 배 태 먼*, 한 규 필*, 이 호 영*, 정 의 윤*, 하 영 호*

Multiresolutional Block-Based Stereo Sequence Coding

Tae-Min Bae*, Kyu-Phil Han*, Ho-Young Lee*, Eui-Yoon Chung*,
Yeong-Ho Ha* *Regular Members*

요 약

본 논문에서는 피라미드 구조를 갖는 다해상도 블럭기반의 스테레오 동영상 부호화기를 제안한다. 제안한 부호화기에서 우영상의 부호화는 기존의 블럭기반의 움직임 예측기법을 사용하고, 좌영상의 부호화에 있어서는 블럭기반의 변이 예측을 이용한다. 또한 좌영상의 시간축 방향의 중복성을 제거하기 위해 우영상의 움직임 정보와 변이 정보로부터 좌영상의 움직임 정보를 유도해 내어 부호화하는 알고리즘을 제안한다. 그리고 제안한 방법에서는 피라미드 구조를 이용하여 부호화시 속도향상을 가져왔으며, 블럭정합 적용시 블럭간의 유사성뿐만 아니라 움직임 벡터와 변이 벡터의 평탄 정도를 동시에 고려하여 각 예측치의 정확성을 높일 수 있었다. 제안한 방법을 연속 스테레오 영상에 적용해 본 결과 블럭정합을 이용하므로 구현이 간단할 뿐만 아니라, 유사성과 평탄성 제약조건을 사용하여 부호화 오차를 줄일 수 있었다.

ABSTRACT

Multiresolutional block-based stereo sequence coding method is presented in this paper. In the suggested codec, the right image sequence is coded with motion estimation and the left image sequence is coded with disparity estimation. Both estimations are based on the multiresolutional block matching method. Since it is possible to calculate rough motion of the left image sequence with disparity and motion of the right image sequence, motion compensation of the left image sequence is performed without motion estimation. Smoothness constraint is added to the criterion of block-matching method for the robust estimation of disparity and motion. Since the proposed codec uses multiresolutional block-based estimation, it is easy to implement and reduces the encoding time. And the experimental results show that the proposed scheme reduces the error of image sequence coding.

* 경북대학교 전자전기공학부
論文番號 : 98009-0106
接受日字 : 1998年 1月 6日

I. 서 론

3차원 가상 현실에 대한 관심이 증가함에 따라 3차원 영상을 경험할 수 있는 시스템이 연구되고 있다. 입체 영상에서는 2차원 영상에서 느낄 수 없는 현실감을 전달할 수 있으므로 스테레오 영상을 이용한 컴퓨터 게임이나 영화가 만들어지고 있으며, 3차원 TV에 대한 연구도 계속되고 있다^{1,2}. 좌·우시각차를 이용하는 스테레오 영상에 의한 입체 영상 시스템의 경우에 물체에 대한 거리감을 느낄 수 있기 때문에 좀더 현실감을 느끼게 할 수 있지만 단일 영상열에 비해 정보량이 두 배로 증가하기 때문에 효율적인 스테레오 영상 부호화에 대한 필요성이 요구된다.

지금까지 알려진 스테레오 영상 부호화의 방법으로는 객체기반(object-based) 방법^{3,6}과 기존의 MPEG과 같은 블럭기반의 움직임 예측방법을 확장한 방법⁷⁻¹⁰이 연구되어왔다. 객체기반 방법은 두 스테레오 영상에서 동일한 3차원 움직임을 가지는 객체를 추출해내고 그 객체에 관한 움직임 정보를 이용하여 영상을 부호화하는데, 스테레오 영상열(stereo image sequence)로부터 좌·우 영상 각각의 움직임(motion) 정보와 좌·우 영상사이의 변이(disparity) 정보를 이용하므로 하나의 영상열(mono image sequence)을 이용하는 경우보다 정확한 3차원 객체 추출과 움직임을 예측할 수 있다. 그러나 객체 기반의 방법은 높은 부호화율을 얻을 수 있지만 알고리즘이 복잡하므로 기존의 블럭기반의 부호화기보다 구조가 복잡하고 객체 추출과 움직임예측에 필요한 최적화가 반복 연산에 의해 수행되어 계산량이 많기 때문에 실시간 부호화기의 구현이 어렵다.

블럭기반의 부호화 방법은 예측시 변이 정보만을 이용하는 경우와 변이 정보와 움직임 정보를 동시에 고려하여 스테레오 영상을 부호화 하는 방법이 있다¹¹. 그러나 변이 정보만을 이용하여 스테레오 영상을 압축하는 경우에 부호화기는 간단해지지만, 변이 정보뿐만 아니라 움직임 정보를 함께 이용하는 경우보다 화질의 저하를 가져온다. 반면 이를 해결하기 위해 움직임 예측과 변이 예측을 동시에 사용하는 경우, 부호화기에 좌·우영상 움직임 예측과 변이 예측을 하기 위해 3개의 예측기가 필요하므로 움직임 예측이나 변이 예측만을 하는 경우의 부호화기보다 복잡해진다.

본 논문에서는 블럭 정합에 기반한 움직임 예측과

변이 예측을 이용하여 부호화를 수행하였다. 우영상의 부호화시 움직임 예측기법을 이용하여 시간축으로의 중복성을 제거하였으며 좌영상의 부호화시 변이 예측기법을 이용하여 좌·우영상간의 중복성을 제거하였다. 또한 좌·우 영상열 사이의 움직임과 변이간에는 상관 관계가 존재하는데 이를 이용하여 우영상의 움직임 정보와 변이 정보에서 좌영상의 움직임 정보를 알아 낼 수 있다. 본 논문에서는 이 유도된 좌영상의 움직임 정보를 이용하여 좌영상에 대해 움직임에 대한 블럭 정합을 수행하지 않고 움직임 보상을 하는 방법을 제안하였다. 그리고 좀더 정확한 예측을 위해서 블럭 정합시 블럭간의 유사성뿐만 아니라 움직임 정보와 변이 정보의 평탄성을 적용하였다. 변이 정보와 움직임 정보 예측시 평탄성을 고려하는 경우 더 나은 결과를 보이지만¹³ 평탄성에 대한 고려를 위해서는 반복 수행에 의해 해를 구해야 하므로 실제 부호화기에서 평탄성을 고려하기는 어렵다. 본 논문에서는 다해상도 영상에서 계층적인 부호화를 수행하는 경우 상위 계층에서의 결과를 이용하여 하위 계층에서 수행하는 부호화를 반복수행의 개념으로 보고 평탄화를 수행하였다. 본 논문에서는 "man"과 "aqua" 스테레오 영상에 대해 실험하였다. 그 결과 제안한 부호화기가 기존의 MPEG기반의 부호화기보다 빠른 부호화를 보여주며, 평탄성을 고려함으로써 좀더 나은 예측 결과를 나타내었다. 또한 변이 정보와 우영상의 움직임 정보를 이용하여 계산된 좌영상의 움직임정보가 영상예측에 적용될 때 예측 오차를 줄임을 알 수 있었다. II장에서는 MPEG기반의 스테레오 부호화기에 대한 일반적인 설명을 하고 III장에서는 제안한 스테레오 부호화기의 전체 구조와 계층적인 움직임 및 변이 예측방법, 변이 정보와 우영상의 움직임 정보에서 좌영상의 움직임 정보를 유도하여 영상 예측에 사용하는 방법에 대해 설명한다. 그리고 마지막으로 IV장에서 제안한 방법의 실험 결과를 나타내었다.

II. 기존의 블럭 정합 기반의 스테레오 영상 부호화

1. 기존의 블럭 정합기반의 스테레오 부호화기의 구조

블럭 정합기반의 스테레오 부호화기에 대해서는 지금까지 많은 연구가 있어 왔다⁷⁻¹⁰. 이들 부호화기는 MPEG-2와 같은 단일 영상 부호화 기법을 기반으로

이를 확장하여 스테레오 부호화에 이용한다. 블록 정합기법의 부호화기는 크게 3가지 방법으로 부호화를 수행하는데, 먼저 블록 기반의 움직임 예측을 수행하여 시간축의 영상열의 중복성을 제거하고, 그 예측 오차에 대해 DCT를 적용하여 공간적인 중복성을 제거한다. 마지막으로 가변 길이 부호화(variable length coding)를 이용하여 부호화를 수행한다^[11]. 스테레오 영상은 그림 1에서와 같이 영상열간의 시간적 중복성과 좌·우 영상사이의 공간적 중복성을 가진다. 그러므로 정확한 움직임과 변이를 추정한다면 부호화율을 극대화시킬 수 있다. 그림 1에서 D_t 와 D_{t+1} 은 t 프레임과 $t+1$ 프레임에서의 변이 정보를 나타내고 V_l 과 V_r 은 각각 좌영상열과 우영상열의 움직임 정보를 나타낸다. 기존의 블록 정합기법의 스테레오 부호화기는 영상열에서 시간적 중복성을 제거하기 위해 움직임 정보인 V_l 과 V_r 을 이용하고, 좌·우영상열간의 공간적 중복성을 제거하기 위해 변이 정보인 D_t 와 D_{t+1} 을 이용한다. 그림 2는 기존의 블록 정합기법의 스테레오 영상 부호화기의 구조이다. 이 때 dv 는 부호화에 필요한 좌·우영상의 변이값이며 R_{in} 과 L_{in} 은 각 부호화기의 좌·우 입력영상을 나타내고 R_{out} 과 L_{out} 은 복호기의 복호된 영상이다. 그러나 기존의 블록 정합기법은 변이 정보의 평탄성이 고려되지 않았을 뿐만 아니라 변이 정보와 움직임 정보간의 상관성도 이용하지 않고 있다. 또한 가장 많은 연산량을 차지하는 예측기가 단일 영상열 부호화기보다 2개 더 필요하다.

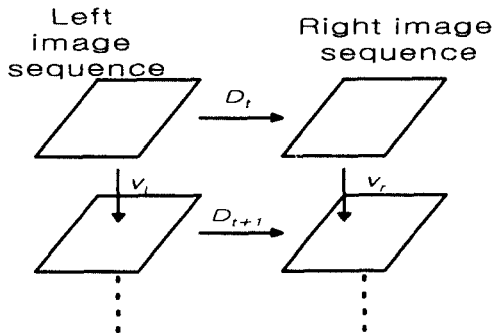


그림 1. 스테레오 영상열간의 관계
Fig. 1 Relation between stereo sequences

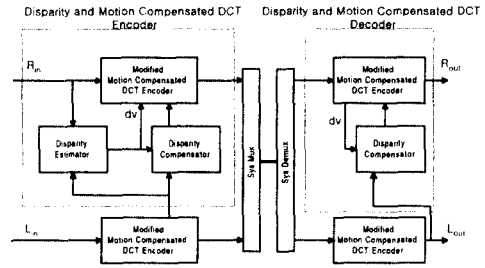


그림 2. 기존의 블록 정합기법의 스테레오 부호화기
Fig. 2 Conventional stereoscopic codec

2. 변이 정보와 움직임 정보에 의한 영상예측

스테레오 영상에서 좌영상을 예측하는 방법에는 우영상과의 변이 정보를 이용하는 경우와 우영상열의 움직임 정보를 이용하는 방법이 있다. 움직임 정보를 이용하는 경우, 예측 오차가 프레임 버퍼에 누적되므로 DCT와 가변 길이 부호화만으로 압축하는 영상내 부호화(intra picture coding)로 복호기의 프레임 버퍼에 저장되어 있는 영상의 오차 전파를 막아야한다^[11]. 그러나 변이 정보를 이용하는 경우는 영상 예측에 사용되는 참조 영상에 이전 영상의 오차가 누적되지 않으므로 위의 경우를 고려할 필요가 없다. 블록 정합 방법은 움직임 탐색시 2차원의 평행한 움직임에 대한 가정하에 탐색을 하므로 확대, 축소, 회전과 같은 복잡한 움직임에 대해서는 해를 구할 수 없으나, 변이 정보는 극상선(epipolar line) 제한조건을 가정한다면 움직임 정보와 달리 일차원적인 거리값으로 나타나므로 블록 정합만으로 원하는 값을 찾을 수 있다. 그러나 움직임 정보는 움직임이 있는 영역에 한정된 반면, 변이 정보는 그 크기가 움직임 정보에 비해 크고 영상 전체에 변이값이 존재하고, 좌·우 영상간에는 카메라 촬영시의 환경에 의해 정합되는 두 영역의 밝기가 차이가 날 수 있어서 변이 정보에 의한 예측 오차는 움직임 정보에 의한 예측 오차보다 일반적으로 크다^[10]. 블록 정합에 의해 변이 정보를 구하는 경우 크게 두 가지 문제가 발생할 수 있다. 먼저, 움직임 정보는 일반적으로 그 벡터의 크기가 작은 데 비해 변이 정보는 크기는 100에서 작게는 1,2정도까지 영상에 따라 크기가 다양하여 변이 예측을 할 때는 움직임 예측보다 더 큰 탐색영역을 필요로 한다. 두 번째로 변이 정보는 평탄성을 가지는데 블록 정합에 의해

탐색을 하면서 평탄성을 고려하기 위해서는 하나의 영상에 대해 반복 탐색을 하여야 한다. 그러나 반복 탐색은 반복 탐색회수에 비례하여 연산량이 증가하여 부호화기에 큰 부담을 주므로 블럭 정합 기법의 전역 탐색을 하는 스테레오 부호화기를 설계하는 경우 평탄성은 정합시 고려되기가 어렵다.

III. 제안한 스테레오 부호화기

1. 제안한 스테레오 부호화기의 구조

그림 3은 제안한 스테레오 영상 부호화기의 블록도이다. 우영상에 대해서는 ME(motion estimation)블록에서 움직임을 추정하고 MC(motion compensation)블록에서 입력 영상과 예측된 영상과의 차영상을 DCT한 후 최종적으로 가변 길이 부호화를 한다. 좌영상의 부호화는 다음과 같다. DE(disparity estimation)블록에서 우영상과의 변이 예측을 한다. ME/DE DEC(motion estimation/disparity estimation decision)블록에서는 변이 정보와 우영상의 움직임 정보로부터 좌영상의 움직임 정보를 유도해낸다. 그리고 나서 유도된 움직임 정보를 이용하여 영상 예측 오차를 계산하고 변이 정보에 의한 영상 예측 오차와 비교를 하여 오차가 작은 정보를 선택하여 영상을 예측한다. 이 때 부호화기에서 복호기로 전송되는 정보는 우영상의 움직임 정보(motion vector)와 좌·우 영상간의 변이 정보(disparity vector), 좌영상의 영상 예측시 변이 정보와 변이 정보에서 유도된 움직임 정보 중 하나를 선택하는 정보(de/me select line), 그리고 좌·우영상의 예측오차이다.

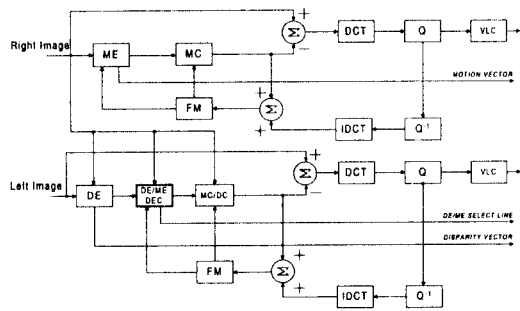


그림 3. 제안한 스테레오 동영상 부호화기의 블록도
Fig. 3 Proposed stereoscopic codec

2. 다해상도 블럭기반의 변이 예측

2.1 움직임과 변이 예측을 위한 계층적 블럭 정합
계층적 탐색의 경우 적정의 계층의 수를 결정해야 하는데 계층수를 늘리면 상위 계층에서의 블럭의 크기가 두 배씩 커지게 되어 탐색에 나쁜 결과를 나타낸다. 실제 계층적 블럭 정합시 적합한 계층수는 3개 정도로 알려져 있다¹⁶⁾.

본 논문에서 영상의 피라미드 계층구조를 얻기 위한 방법은

$$I_{x,y}^L = \frac{1}{4} \sum_{i=0}^L \sum_{j=0}^L I_{2x+i, 2y+j}^{L-1} \quad (1)$$

와 같다. 여기서 L 은 계층의 레벨을 나타내며 레벨이 클수록 상위 계층을 나타내고 x, y 는 영상에서의 화소의 위치이다. 일반적으로 피라미드 계층구조를 얻기 위해서는 영상을 저역통과필터에 통과시킨 후 행렬방향으로 간축(down sampling)을 행하게 된다. 본 논문에서는 실제 구현이 용이한 평균필터(mean filter)를 사용하였다.

각 계층에서 사용되는 블럭의 크기를 설정할 때 계층별로 다른 크기의 블럭을 이용하는 방법과 같은 크기의 블럭을 사용하는 방법이 있을 수 있다. 두 방법에 대한 비교를 보면 계층별로 같은 크기의 블럭을 사용하는 것이 더 향상된 결과를 보여주었다¹⁶⁾. 계층적인 블럭 정합의 경우 그 하위 계층에서의 탐색은 상위계층에서의 탐색에 큰 영향을 받게 된다. 즉 상위 계층에서는 낮은 해상도에서 넓은 영역을 탐색하며 하위 계층에서는 상위 계층에서 찾은 정합 결과를 좀 더 높은 해상도에서 자세히 탐색하게 되는데, 상위 계층에서 잘못된 결과를 찾게 되면 그 하위 계층에서의 탐색으로는 지역적인 해(local solution)만을 구하게 된다. 그리고 각 계층마다 같은 크기의 블럭을 사용하므로 상위 계층의 하나의 블럭은 하위 계층에 4개의 블럭에 대응되어 하위 계층의 4개 블럭은 비슷한 변이 값을 가지게 된다. 그러나 실제로 상위 계층에서 찾은 결과가 잘못될 수도 있으며 4개의 블럭이 매우 다른 변이 값을 가질 수도 있다. 이런 경우 하위 계층에서 탐색 영역을 더 크게 하여 재 탐색하는 방법을 생각할 수 있지만¹²⁾ 이런 경우 탐색에 걸리는 시간은 영상에 따라 다르게 된다. 실제로 부호화기를 설계할 때는 탐색에 사용되는 최대 시간을

고려해야 하므로 이 방법은 전역 탐색에 비해 더 많은 시간이 걸릴 수 있다. 다른 방법으로는 주위 블럭의 결과를 이용하는 방법을 생각할 수 있는데, 사용되는 블럭의 크기가 8×8 로 작기 때문에 주위 블럭들과 같은 움직임이나 변이를 가질 가능성은 매우 크다. 예를 들면 그림 4에서와 같이 영역이 분할되는 경우를 가정하면 단순히 상위 계층에서의 탐색 결과를 하위 계층에서 이용하는 방법으로는 올바른 해를 구할 수 없다. $L+1$ 계층에서 블럭 D_4 의 정합 결과를 L 계층의 S_1, S_2, S_3, S_4 블럭의 탐색의 초기값으로 사용되게 되면 S_4 블럭의 경우 D_8 의 움직임정보를 초기값으로 사용하는 것이 더 타당하다. 이와 같은 경우 하위 계층에서 탐색의 초기값으로 관계되는 상위 계층의 블럭의 탐색 결과만을 이용하기 때문에 생기는 문제라고 볼 수 있다. 즉 탐색의 초기치로 사용할 수 있는 값들을 모두 초기값 후보로 하여 가장 초기치로 적합한 값을 선택함으로써 이 문제를 해결할 수 있다. 초기치의 후보로 사용할 수 있는 값으로는 상위 계층의 관계되는 블럭의 탐색 결과와 그 주위 블럭의 탐색 결과값이 있다.

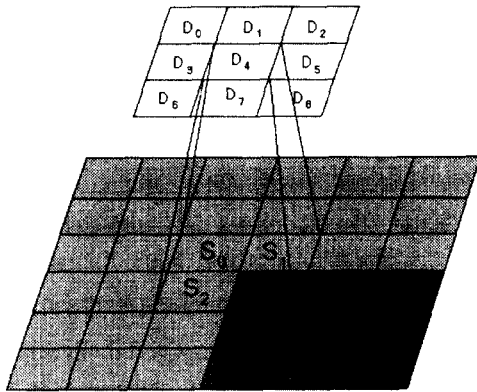


그림 4. 계층적 블럭 정합
Fig. 4 Hierarchical block matching

본 논문에서는 상위계층에서의 탐색결과를 최대한으로 사용하기 위해 다음과 같은 방법을 적용하였다. 현재 블럭에 대응되는 상위 계층에서의 블럭과 상위 계층에서의 주위 8개의 블럭의 변이 예측값을 초기값의 후보로 하여 전체 9개의 후보 벡터에 대한 블럭 정합

정도를 구하여 가장 작은 오차를 가지는 값을 초기값으로 선택하였다. 그림 5에서 보면 $L+1$ 계층에서의 블럭이 L 계층에서 4개의 블럭으로 나뉘어진다. 이때 L 계층에서 4개의 블럭의 변이 초기치로 $L+1$ 계층에서 대응되는 블럭의 변이 값과 주위 블럭인 $D_0, D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8$ 의 변이 값을 이용한다.

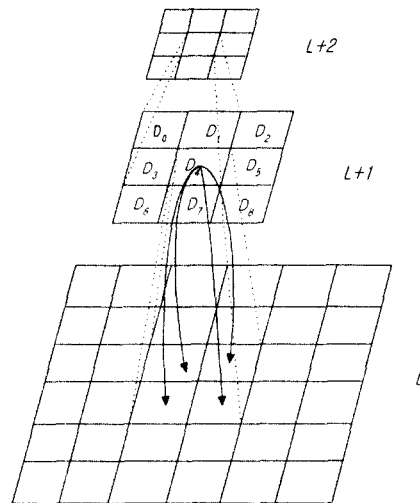


그림 5. 피라미드 영상구조
Fig. 5 Pyramid image structure

2.2 전역탐색과 계층적 탐색의 계산량 비교

계층적 탐색을 수행하는 경우 상위 계층에서의 영상은 그 하위 계층의 영상에 비해 탐색영역이 크게 줄어든다. 예를 들어 계층의 수가 1인 경우, 각 계층에서의 탐색 영역을 $s \sim s$ 로 설정하면 전체 영상에서는 $2(2^l - 1)s + 1$ 의 탐색영역을 가지는 것과 같다. 이러한 탐색영역의 감소는 영상에서의 정합 오류나 중복 정합을 크게 줄일 수 있으며, 계산 시간의 감소 효과를 가져온다. $N \times N$ 크기의 블록으로 정합을 수행하는 경우 한 위치에서의 정합에는 N^2 의 픽셀, $N^2 - 1$ 의 덧셈 그리고 N^2 의 곱셈이 필요하므로 총 $3N^2 - 1$ 의 연산이 필요하다. $a \sim a$ 의 탐색 영역에 대해 움직임 예측을 수행하는 경우 전역탐색으로 탐색해야할 위치의 수는 $(2a + 1)^2$ 개이다. 그러므로 전역 탐색에

의한 움직임 예측시 연산량은 다음과 같다.

$$C_{F.M.} = (2a+1)^2 \times (3N^2 - 1) \quad (2)$$

제안한 계층적 블록 정합은 l 개의 계층을 가지는 경우 각 계층별로 $-\frac{2a}{2^l-1} \sim \frac{2a}{2^l-1}$ 의 영역을 탐색하고, 초기치를 결정 하기 위해 상위 계층의 초기치 후보값 8개를 탐색한다. 그리고 상위 계층의 블록 정합을 위한 연산이 하위 계층에서 4개의 블록의 정합에 사용되는 것을 고려하면 움직임 예측에 필요한 연산량은

$$C_{P.M.} = \left\{ \frac{4}{3} \times (1-4^{1-l}) \times \left(\frac{2a}{2^l-1} + 1 \right)^2 + 8 \times (l-1) \right\} \times (3N^2 - 1) \quad (3)$$

로써, 블록정합시 탐색해야할 지점의 수가 줄어들어 전체 연산량이 줄어드는 것을 알 수 있다. 변이 예측시의 탐색영역이 $-b \sim b$ 일때 전역 탐색에 의한 연산량은

$$C_{F.D.} = (2b+1) \times (3N^2 - 1) \quad (4)$$

이고, 계층적 탐색에 의한 연산량은

$$C_{P.D.} = \left\{ \frac{4}{3} \times (1-4^{1-l}) \times \left(\frac{2b}{2^l-1} + 1 \right) + 8 \times (l-1) \right\} \times (3N^2 - 1) \quad (5)$$

와 같다. 3개의 계층을 가지는 계층구조의 탐색과 전역탐색을 비교하기 위해 탐색영역을 $-16 \sim 16$ 으로 가정하면 움직임 정보의 탐색시 8배정도 전역탐색이 연산량이 많고, 변이 정보의 탐색시에는 약 1.5배의 연산량을 나타내었다. 그러나 변이 정보는 움직임 정보에 비해 그 크기가 크므로 변이의 탐색영역을 $-50 \sim 50$ 으로 설정하면 전역 탐색에 의한 연산량이 계층적인 정합방법에 비해 3배 정도 크고 탐색영역이 증가하면 연산량의 차이는 더욱 커진다. 변이 정보와 움직임 정보를 모두 탐색하는데 필요한 연산량을 계산하면 표 1과 같이 약 6배 정도 전역 탐색이 계층적 탐색에 비해 연산량이 많다. 즉 전역 탐색에 의해 움직임 추정만을 수행하는데 필요한 연산량이 계층적 탐색에

의해 변이와 움직임을 모두 탐색하는데 소요되는 연산량보다 약 5배 정도 많기 때문에 전역 탐색으로 움직임 추정을 하는데 필요한 시간으로 계층적 탐색을 하면 움직임 추정과 변이 추정을 모두 예측할 수 있다. 그러므로 계층적 블록정합을 수행하면, 하나의 예측기를 사용하여 움직임과 변이를 시 분할하여 탐색할 수 있으므로 부호화기를 구현 할 때 부호화기를 간단히 할 수 있다.

표 1. 계층적 탐색과 전역 탐색시 계산량 비교: 변이 탐색 영역은 $-50 \sim 50$, 움직임 탐색 영역은 $-16 \sim 16$ (단위: 사칙연산의 횟수)

Table 1. The comparison of computational load between Hierarchical and full search for one block. $-50 \sim 50$ for disparity search, $-16 \sim 16$ for motion search ; the number of addition or multiplication is counted

method	disparity	motion	disparity and motion
hierarchical search	6706	10467	1717303522
full search	19291	84231	103522

3. 변이 정보와 우영상의 움직임 정보를 이용한 좌영상의 움직임 정보 예측

3.1 변이 정보와 움직임 정보의 관계

평행하게 정렬된 스테레오 카메라에서 실제 물체가 이동할 때 우영상에서의 움직임 정보와 좌영상의 움직임 정보와의 관계를 유도해 낼 수 있다. 그림 6에서 두 카메라의 중심간 거리는 B 이고, 투영된 영상의 크기는 $2S \times 2S$, 그리고 f 는 카메라의 초점거리이다. t 시간에서 a_t 는 물체의 위치, l_t 는 좌영상에서의 대응되는 위치, r_t 는 우영상에서의 대응되는 위치를, 그리고 Z_t 는 카메라 초점거리에서 물체까지의 거리이다. 여기에서 변이 정보와 카메라에서 물체의 거리를 구하면 다음과 같다.

$$\begin{aligned} Z_{t+1} : Z_t &= (l_t - r_t) : (l_{t+1} - r_{t+1}) \\ &= -D_t : -D_{t+1} \end{aligned} \quad (6)$$

$l_{t+1} - l_t = v_l$, $r_{t+1} - r_t = v_r$, $D_t = r_t - l_t$ 에서 영상의 움직임 벡터를 좌영상의 움직임 벡터와 변이값으로 정리하면

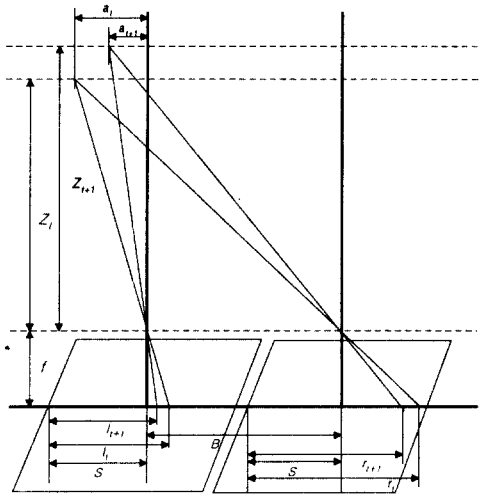


그림 6. 스테레오 영상에서 움직임과 변이와의 관계
Fig. 6 Relation between motion and disparity information in stereo image

$$v_l = D_t - D_{t+1} + v_r \quad (7)$$

과 같다. 즉 좌영상의 움직임 정보 v_l 을 우영상의 움직임 정보 v_r 와 두 프레임의 변이 정보 D_t 와 D_{t+1} 에서 구할 수 있음을 알 수 있다. 그러나 위의 결과는 화소단위에서는 그대로 적용할 수 있지만 블럭단위 정합에서는 그대로 적용할 수가 없다. 블럭단위 정합을 이용한 움직임 정보의 예측방법은 이차원 평행한 움직임만이 정확하게 예측될 수 있다. 그러므로 실제 물체에서 이차원 평행한 움직임(2D translational motion)만이 있다고 가정하면

$$Z_t = Z_{t+1} \quad (8)$$

이 성립하고 식 (7)은 다음과 같이 간단하게 표현될 수가 있다.

$$v_l = v_r \quad (9)$$

즉 우영상의 움직임정보를 정합되는 좌영상에 그대로 적용할 수가 있다. 이를 바탕으로 좌영상의 k 번째 블럭의 움직임 정보를 계산하는 식은 다음과 같다.

$$V_l(k) = V_r(k + \text{SIGN}(D(k)) \times \text{ROUND}[\frac{-D(k) + B/2}{B}]) \quad (10)$$

이때 ROUND 함수는 부동 소수점을 정수 값으로 변환하는 함수이고 SIGN 함수는 입력값의 $+/-$ 부호를 반환하는 함수이다. 정합 블럭의 크기가 $B \times B$ 일 때 좌영상의 k 번째 블럭의 움직임 벡터 $V_l(k)$ 는 k 번째 블럭의 변이 정보 $D(k)$ 를 이용하여 우영상의 $(k + \text{SIGN}(D(k)) \times \text{ROUND}[\frac{-D(k) + B/2}{B}])$ 번째 블럭의 움직임 정보를 이용하게 된다.

본 논문에서는 영상을 예측할 때 블럭단위로 변이 정보에 의한 예측 오차와 움직임 정보에 의한 오차를 비교해서 작은 값을 가지는 것을 사용하므로 매 블럭마다 움직임정보와 변이정보중 어느 것을 사용할 지를 알리는 1-bit가 추가된다. 그러나 기존의 움직임과 변이 정보를 모두 전송하는 시스템과 비교할 때 제안한 방식은 좌영상의 움직임 정보를 전송할 필요가 없으므로 정보량은 감소하게 된다.

3.2 변이 정보와 움직임 정보의 평탄성의 고려

우영상의 움직임 정보와 변이 정보로부터 계산된 좌영상의 움직임 정보는 정확한 움직임 정보가 아니므로 우영상에서 인접한 두 블럭의 움직임 벡터가 많이 다르다면 위의 결과를 적용하는 경우 큰 오차를 유발할 수 있다. 이러한 경우를 피하기 위해서 변이 예측과 움직임 예측시 평탄한 정도를 정합 정도에 추가함으로써 평탄하게 변하는 움직임 정보와 변이 정보를 구할 수 있다. 또한 평탄성을 고려하여 반복 탐색 하면 탐색 결과에 대해 이완 처리를 하여 잡음에 의한 잘못된 결과를 제거 할 수 있다. 움직임 정보나 변이 정보를 반복 수행으로 찾아내는 방법들은 벡터의 평탄한 정도를 고려함으로써 좀더 나은 탐색 결과를 얻어내고 있으며^[13-15] 블럭 정합기법에 변이 벡터의 평탄화를 적용하기 위해서 영상에 대하여 반복 정합 수행을 하는 방법들도 있다.^[13] 그런데 계층적인 블럭 정합 기법에서는 각 계층별로 블럭 정합이 수행되므로 반복 수행을 하지 않고 영상에 대해 계층수 만큼의 반복 수행을 하는 효과를 얻을 수 있다. 블럭 정합시에는 현재 탐색하고 있는 블럭의 주위 블럭의 탐색값을 알 수가 없으므로 그림 5에서와 같이 상위 계층에 있는 주위 블럭의 탐색 결과를 이용하여 평탄정도를 계산한다. 식(11)은 부호화되고 있는 블럭의 탐색값과

주위 블럭의 탐색값과의 평탄한 정도(smoothness)를 나타낸다.

$$S = \sum_{n=0}^7 (D - D_n)^2 \quad (11)$$

여기서 D 는 탐색중의 블럭의 예측값이고 D_n 은 상위 계층의 주위 블럭의 탐색값이다. 블럭의 유사성과 탐색값의 평탄화 정도를 모두 고려한 블럭 오차함수는 다음과 같다.

$$E = \sum_{(x,y) \in \omega} (I_r(x,y) - I_l(x,y+D))^2 + \alpha \times \sum_{n=0}^7 (D - D_n)^2 \quad (12)$$

여기서 ω 는 블럭내 영역, D 는 좌·우 블럭간의 탐색값을 나타내며 α 는 평탄정도를 고려하는 가중치로 영상의 잡음이 큰 경우 값을 크게 할수록 좋은 결과를 가져온다.

IV. 실험 및 고찰

384×384 크기를 가지는 25프레임의 'man' 영상과 640×512크기의 25프레임의 'aqua' 스테레오 영상이 실험에 사용되었다. 실험에 사용된 스테레오 영상은 극상선(epipolar line) 제한 조건을 만족시키므로 8×8 단위의 블럭에 대해 일차원 탐색만을 수행하여 변이 예측을 하였다. 블럭 정합은 화소단위(integer pel)로 하였으며 연속된 5 프레임씩에 대하여 실험을 하였다. 단, 변이 예측과 움직임 예측시 MPEG에서 사용되는 양방향 예측은 하지 않았으며 영상내 부호화를 하지 않고 두 번째 프레임부터 예측 부호화를 하였다. 표 2와 3은 'man'과 'aqua' 스테레오 동영상에 대해 각 예측 방법에 의해 얻어진 복원 영상의 평균 PSNR이다. case 1은 피라미드 구조를 사용하지 않는 전역 탐색, case 2는 계층적 탐색을 이용한 변이예측, case 3은 계층적 탐색으로 찾은 우영상의 움직임 정보와 변이 정보에서 좌영상의 움직임 정보를 계산하여 영상 예측을 한 경우, 그리고 case 4는 case 3의 방법에서 탐색시 평탄성을 고려하였을 때의 결과이다. 제안한 알고리즘에 의해 복원된 영상의 PSNR은 변이 정보만을 이용한 경우보다 평균 1.5dB에서 3dB정도 향상되었다. 실험에서 사용한 계층적 탐색이 전역탐색에 비해 계산량이 적으면서도 예측 결과가 유사하므로 보

다 유용한 방법임을 알 수 있다. 또한 표 2, 표 3에서 보듯이 평탄성을 고려한 case 4의 결과가 case 3보다 나은 결과를 보였다.

표 2. 복원된 'man' 좌영상의 평균 PSNR

Table 2. Average PSNR of the left man sequence(case 1: disparity estimation with full search, case 2: disparity estimation with hierarchical search, case 3: proposed method without smoothness criteria, case 4: proposed method with smoothness criteria)

searching method	PSNR
case 1	40.33
case 2	40.09
case 3	41.53
case 4	41.91

표 3. 복원된 'aqua' 좌영상의 평균 PSNR

Table 3. Average PSNR of the left aqua sequence(case 1: disparity estimation with full search, case 2: disparity estimation with hierarchical search, case 3: proposed method without smoothness criteria, case 4: proposed method with smoothness constraint)

searching method	PSNR
case 1	33.35
case 2	33.36
case 3	38.36
case 4	38.38

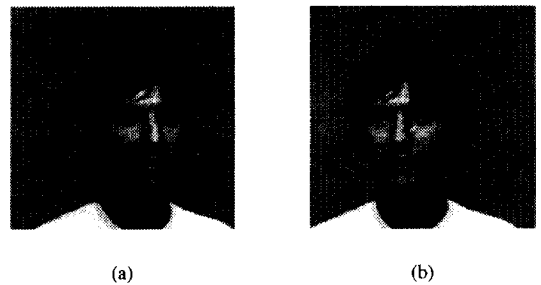


그림 7. 'man' 스테레오 영상 (a) 첫번째 프레임 좌영상 (b) 첫번째 프레임 우영상

Fig. 7. 'man' stereo sequence. (a) left image of first frame. (b) right image of first frame

제안한 알고리즘은 변이 정보와 우영상의 움직임 정보에서 좌영상의 움직임 정보를 유도하므로, 여기서 얻어진 움직임 정보를 이용하여 변이 정보만으로는 예측이 불가능한 영역까지도 고려할 수 있다. 예를 들어, 원영상인 그림 8에서 좌영상의 좌측 하단 모서리 배경은 우영상에서 존재하지 않으므로 변이 정보만으로는 예측할 수 없지만, 제안한 알고리즘은 시간적 상관성을 이용하여 오차를 줄이는 것을 그림 9에서 확인할 수 있다. 제안한 알고리즘을 움직임 예측만을 한 경우와 비교를 하기 위해 그림 14에서 'aqua' 영상에 대하여 움직임 예측에 의해 추정된 영상을 그림 10에서 'man' 영상에 대하여 움직임 예측에 의해 추정된 영상과 원영상과의 차영상을 나타내었다. 그림 15와

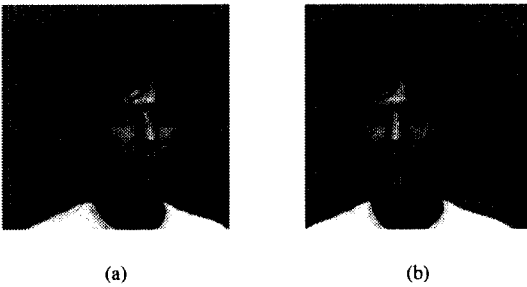


그림 8. 'man' 스테레오 영상 (a) 두번째 프레임 좌영상 (b) 두번째 프레임 우영상
Fig. 8 'man' stereo sequence. (a) left image of second frame. (b) right image of second frame

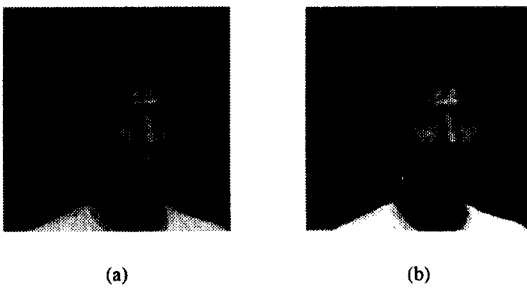


그림 9. 두 번째 프레임의 예측된 좌영상 (a) 제안한 알고리즘으로 예측된 영상 (b) 변이 정보를 이용하여 예측된 영상
Fig. 9 Estimated image of second left frame by (a) proposed method and (b) disparity estimation.

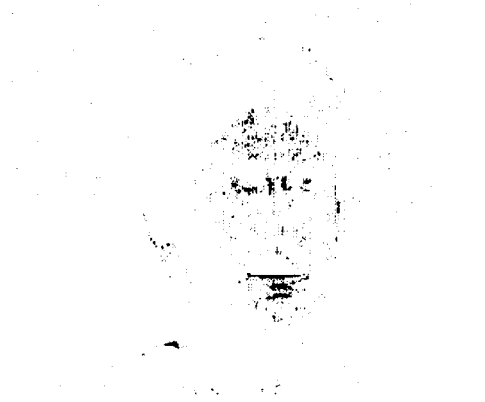


그림 10. 'man' 영상 : 움직임 정보를 이용하여 예측시의 오차 영상
Fig. 10 difference image between original and motion compensated image : 'man' sequence

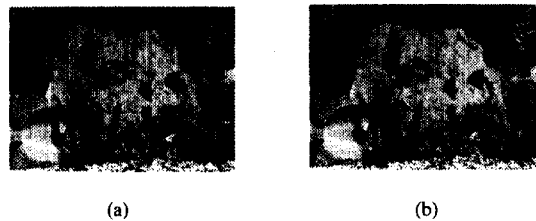


그림 11. 'aqua' 영상 (a) 첫번째 프레임 좌영상 (b) 첫번째 프레임 우영상
Fig. 11 'aqua' stereo sequence. (a) left image of the first frame (b) right image of the first frame

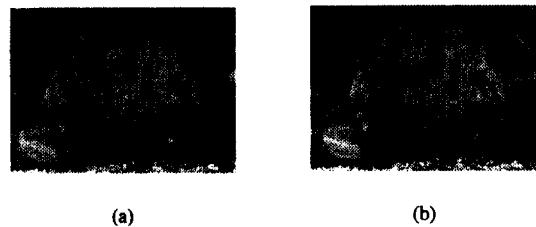


그림 12. 'aqua' 영상. (a) 두번째 프레임 좌영상. (b) 두번째 프레임 우영상.
Fig. 12 'aqua' stereo sequence. (a) left image of the second frame. (b) right image of the second frame.



(a)



(b)

그림 13. 두 번째 프레임의 예측된 좌영상 (a) 제안한 알고리즘으로 예측된 영상 (b) 변이 정보를 이용하여 예측된 영상

Fig. 13 Estimated image of second left frame by (a) proposed method and (b) disparity estimation.

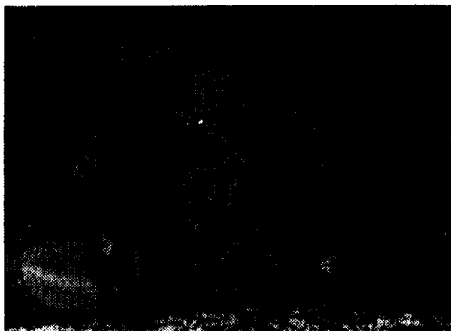


그림 14. 'aqua'영상 : 움직임 정보를 이용한 예측 영상
Fig. 14 difference image between original and motion compensated image : 'aqua' sequence

그림 16은 'man'영상의 25프레임에 대한 각 프레임 별 실험 결과로 그림 15는 변이 예측만을 이용한 경우와 제안한 알고리즘에 의한 결과를 비교하였으며 그림 16은 평탄성 고려 여부에 따른 결과이다. 그림 15에서 알 수 있는 바와 같이 제안한 알고리즘에서 좌영상에 대해 움직임 보상을 하는 경우 프레임 비퍼에 저장된 영상을 영상예측에 사용하므로 오차가 누적된다. 이 때문에 제안한 알고리즘에서의 움직임 정보를 부호화에 적용하여 변이 정보를 이용한 경우보다 높은 PSNR을 나타내기 위해서는 좌영상 또한 영상내 부호화를 하여야 움직임 정보에 의한 부호화의 효과를 얻을 수 있으며 그렇지 않다면 결국 변이 정보에 의한 부호화만이 좌영상의 PSNR에 영향을 미치게 된다. 가시도를 높이기 위해 그림 10은 문턱치값(threshold) 이상의 오차만을 나타내었으며 다른 영상은 명암을 10~20% 올렸다.

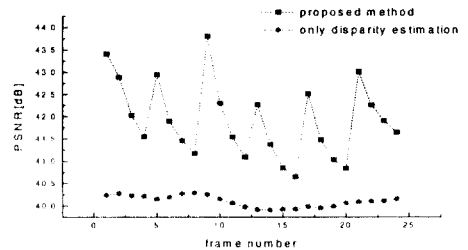


그림 15. 'man' 복원영상의 PSNR : 제안한 방법과 변이 추정방법의 비교

Fig. 15 PSNR of decoded image of 'man' sequence: comparison of proposed method and disparity estimation.

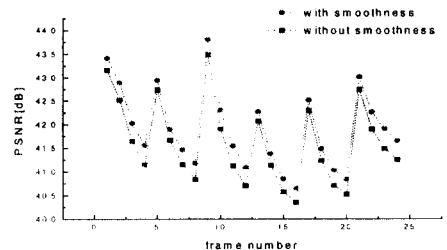


그림 16. 평탄성 제약조건에 대한 'man' 좌영상의 PSNR 비교
Fig. 16 The comparison of PSNR for the left 'man' sequence according to smoothness constraint.

V. 결 론

본 논문에서는 스테레오 동영상 부호화에 피라미드 구조의 다해상도 접근을 사용한 방법을 제안하였다. 제안한 방법은 스테레오 영상 부호화시 변이 정보를 이용하여 좌·우 영상 사이의 공간적 중복성을 제거 하였으며, 변이 정보만을 예측하면서도 움직임 정보를 스테레오 카메라의 기하학적 구조에서 계산함으로써 시간적인 상관성도 고려하여 변이만을 사용하는 경우 보다 높은 PSNR을 얻을 수 있었다. 계층적인 블럭 정합 방법을 이용하여 빠른 부호화가 가능하였고 1개의 예측기를 사용하여 전역탐색에 의한 움직임 예측에 필요한 연산량의 28%만으로 움직임 예측과 변이 예측이 가능함을 보여 주었다. 또한 계층적 블럭 정합이 동일한 영상에 대해 반복적인 탐색을 하므로 전역 탐색의 경우 고려할 수 없었던 변이 정보와 움직임 정보의 평탄성을 고려할 수 있었다. 본 논문에서 제안한 부호화기는 블럭 정합을 이용함으로써 MPEG-2 기법의 기존 방법과도 쉽게 호환될 수 있는 장점이 있다.

참 고 문 헌

1. F. Chassaing, B. Choquet and D. Pele, "A stereoscopic television system (3D-TV) and compatible transmission on a MAC channel (3D-MAC)," *Signal Processing Image Communication*, vol. 4, pp. 33-43, 1991.
2. R. Horst, "A digital codec for 3D-TV transmission," *Signal Processing of HDTV, IV*, pp. 489-495, 1993.
3. D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "Object-based coding of stereo image sequences using joint 3-D motion/disparity compensation," *IEEE Trans. on C.S.V.T* vol. 7, pp. 312-327, 1997.
4. S. Panis and M. Ziegler, "Object based coding using motion and stereo information," in *Picture Coding Symp. (PCS '94)*, pp. 308-312, Sep. 1994.
5. D. Tzovaras, N. Grammalidis, and M. G. Strintzis, "3-D motion/disparity segmentation for object-based image sequence coding," *Opt. Eng., Speial Issue on Visual Commun. Image Processing*, vol. 35 pp. 308-312, Jan. 1996.
6. N. Grammalidis, S. Malassiotis, D. Tzovaras, and M. G. Strintzis, "Stereo image sequence coding based on three-dimensional motion estimation and compensation," *Signal Processing Image Communication*, pp. 129-145, Apr. 1994.
7. 권동현, 장현식, 김진웅, 이명호, 김영식 "MPEG-2 MVP를 이용한 입체 영상 부호화기," *대한전자공학회 추계종합학술대회 논문집*, 제19권, B편, 제2호, pp. 1191-1194, 1996.
8. M. G. Perkins, "Data compression of stereopairs," *IEEE Trans. on Communications*, vol. 40, no. 4, pp. 684-696, 1992.
9. Y. J. Song, "Improved disparity estimation algorithm with MPEG-2's scalability for stereoscopic sequence," *IEEE Trans. on Consumer Electronics*, vol. 42, no. 3, pp. 306-311, 1996.
10. A. Puri, R. V. Kollarits, and B.G. Haskell, "Compression of stereoscopic video using MPEG-2," *Critical Reviews conference on Standards and Common Interfaces for Video Information Systems*, published by SPIE, Oct. 1995.
11. Test Model Editing Committee, "MPEG-2 video test model 5," *ISO/IEC JTC1/SC29/WG11 Doc. N0400*, Apr. 1993.
12. M. Accame, Francesco G.B. De Natale, and Daniele D. Giusto, "Hierarchical block matching for disparity estimation in stereo sequence," *ICIP*, pp. 374-377, 1995.
13. Woontack Woo and Antonio Ortega, "Stereo image compression with disparity compression using MRF model," *VCIP/SPIE*, vol. 2727, pp. 28-39, 1996.
14. S. Malassiotis, N. Nikolaidis, M. Strintzis, and Pittas, "A smoothing algorithm with reliability constraints for motion vector field coding," *Proc. Internat. Cong. Communication and Control*, Rhodes, pp. 177-182, June 1993.
15. Jin Liu and Robert Skerjanc, "Stereo and motion correspondence in a sequence of stereo images," *Signal Processing Image Communication*, vol. 5, pp. 305-318, 1993.
16. D. Tzovaras, M. G. Strintzis, and H. Sahinoglou,

"Evaluation of multiresolution block matching techniques for motion and disparity estimation,"
Signal Processing Image Communication, vol. 6,
pp. 59-67, 1994.



배 태 면(Tae-Min Bae) 정회원
1974년 1월 4일생
1996년 2월 : 경북대학교 공과대학
전자공학과 졸업(공학
학사)
1996년 2월 : 경북대학교 대학원 전
자공학과 졸업(공학
석사)

1998년 3월~현재 : 경북대학교 대학원 전자공학과 박사과정

※주관심분야 : 스테레오 정합, 영상분할, 객체기반 스테레오 동영상 부호화등.

한 규 필(Kyu-Phil Han) 정회원
1968년 10월 24일생
한국통신학회 제22권 4호 참조
현재 : 경북대학교 대학원 전자공학과 박사과정

이 호 영(Ho-Young Lee) 정회원
1972년 1월 3일생
한국통신학회 제22권 4호 참조
현재 : 경북대학교 대학원 전자공학과 박사과정

정 의 윤(Eui-Yoon Chung) 정회원
1969년 6월 16일생
한국통신학회 제22권 9호 참조
현재 : 경북대학교 대학원 전자공학과 박사과정

하 영 호(Yeong-Ho Ha) 정회원
한국통신학회 제22권 4호 참조
현재 : 경북대학교 전자전기공학부 교수