

객체 분할 기법을 이용한 다시점 영상 부호화에서의 예측 모드 선택 기법

(A Mode Selection Algorithm using Scene Segmentation for Multi-view Video Coding)

이 서 영 [†] 신 광 무 [†] 정 기 동 ^{**}
(Seoyoung Lee) (Kwangmu Shin) (Kidong Chung)

요 약 최근 멀티미디어 기술의 발달과 더불어 3차원 영상에 대한 연구가 활발하게 이루어지고 있다. 이 중 다시점 영상은 사실감 넘치는 화면을 사용자에게 제공하지만, 대역폭의 급격한 증가는 풀어야 할 주요 문제이다. 본 논문은 부호화 과정의 복잡도와 시간 소모를 줄일 수 있는 빠른 예측모드 결정 알고리즘을 제안한다. 이것은 빠르게 움직이는 전경 객체를 효과적으로 구분할 수 있는 객체 분할을 기반으로 한다. 빠른 움직임의 가진 전경 객체가 시점 방향 예측 모드로 부호화 될 가능성이 더 높기 때문에 움직임 보상 과정을 사전에 제한할 수 있다. 제안한 기법을 적용한 결과, 기존의 부호화 과정과 비교하여 화질의 큰 저하 없이 평균 45% 연산량이 감소하였다.

키워드 : 다시점 영상 부호화, 예측 모드 선택, 계산 복잡도 감소

Abstract With the growing demand for multimedia services and advances in display technology, new applications for 3-D scene communication have emerged. While multi-view video of these emerging applications may provide users with more realistic scene experience, drastic increase in the bandwidth is a major problem to solve.

In this paper, we propose a fast prediction mode decision algorithm which can significantly reduce complexity and time consumption of the encoding process. This is based on the object segmentation, which can effectively identify the fast moving foreground object. As the foreground object with fast motion is more likely to be encoded in the view directional prediction mode, we can properly limit the motion compensated coding for a case in point. As a result, time savings of the proposed algorithm was up to average 45% without much loss in the quality of the image sequence.

Key words : Multi-view Video Coding, Fast Mode Selection

1. 서 론

최근멀티미디어 수요의 증가와 디스플레이 장치 기술

의 발전과 함께 3차원 영상에 대한 연구가 활발하게 진행되고 있다. 3차원 영상은 깊이와 시차 정보를 이용하여 화면에 자연스러운 입체감을 부여한다. 이러한 기술의 응용 형태로는 입체감 있는 화면을 다각도로 시청할 수 있는 3DTV와 시점 변환을 통해 영상의 다양한 시점을 선택적으로 볼 수 있는 자유 시점 TV(free view-point TV)[1], 보다 넓은 시야의 화면을 제공하는 파노라마(panorama) 영상 등이 있다.

다시점 영상은 2대 이상의 카메라로 획득한 영상이며, 관찰자의 시점이 고정되는 기존의 스테레오(stereo) 영상에 비해 피로감 없는 자연스러운 입체감을 제공한다. 하지만 이러한 특징에도 불구하고 데이터의 양이 시점에 비례하여 증가하며 카메라 간 영상의 차이로 인해 예측 과정이 복잡해지는 단점이 있다. 따라서 이를 효율적으로 압축, 전송하는 기술에 대한 연구 및 표준화 작

· 이 논문은 제35회 추계학술대회에서 객체 분할 기법을 이용한 다시점 영상 부호화에서의 예측 모드 결정 방법의 제목으로 발표된 논문을 확장한 것임

[†] 학생회원 : 부산대학교 컴퓨터공학과
memoiry@pusan.ac.kr
sin@pusan.ac.kr

^{**} 종신회원 : 부산대학교 컴퓨터공학과 교수
kdchungm@melon.cs.pusan.ac.kr

논문접수 : 2009년 1월 21일
심사완료 : 2009년 4월 13일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

업이 진행 중에 있다. MPEG(Moving Picture Experts Group)은 3DAV(3D Audio/Visual)라는 명칭으로 2002년부터 3차원 영상 부호화의 표준화 작업을 수행하였으며, 현재 JVT(Joint Video Team)의 주도 하에 다시점 영상 압축 기술의 표준화 작업이 진행되고 있다[2].

다시점 영상을 효율적으로 압축하기 위한 방법으로는 크게 예측 부호화 기법(predictive coding)과 대역 분할 부호화 기법(subband coding)이 있다. 예측 부호화 기법은 이전에 부호화된 화면(picture)을 참조 화면(reference picture)으로 사용하여 현재 화면을 예측하며, 현재 표준화 작업은 이와 같은 기법에 기반을 둔 H.264/AVC를 중심으로 이루어지고 있다. 또한 다시점 영상을 효율적으로 부호화하기 위해 기존 H.264에서 이용하는 시·공간적 상관성뿐만 아니라 시점 간 상관성을 활용한다. 이는 각각의 카메라에서 획득한 동시간대의 영상을 참조 화면으로 사용하여, 기존의 방법에 비해 예측의 효율성이 높다[3]. 하지만 기존의 방법에 비해 예측 과정이 복잡하고 더 많은 화면을 참조해야 하므로 연산량이 크게 증가하기 때문에, 부호화 속도를 향상시키기 위해서는 이를 효과적으로 수행할 필요가 있다.

본 논문은 기존의 움직임 보상 예측 과정(motion compensated prediction)과 시점 간 예측(disparity compensated prediction)을 보다 효율적으로 수행하는 기법을 제안한다. 이는 화면상의 빠르게 움직이는 객체를 구분해내고 객체의 특성에 맞게 예측 모드를 미리 결정함으로써 가능하다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구, 3장에서는 제안하는 기법을 설명하며, 4장에서는 실험 결과를 기술한다. 마지막으로 5장에서 결론으로 끝을 맺는다.

2. 관련 연구

다수의 카메라로 획득한 다시점 영상의 특성상, 시점이 증가함에 따라 처리해야 할 데이터의 양이 증가한다. 시간적으로 인접한 참조 화면에 대한 예측만을 수행하던 기존의 영상 부호화 과정과는 달리, 다시점 영상 부호화는 시점 간 예측 과정을 같이 수행하기 때문에 예측 과정이 복잡해지고 이로 인해 연산량이 크게 증가한다. 따라서 예측 과정에서 소요되는 연산량을 감소시키기 위한 여러 방법들이 연구되었다.

Li 등은[4] 시간 참조 화면에서 현재 블록과 동일한 위치에 있는 참조 블록의 모드 정보를 이용하여 현재 블록의 부호화 방향을 결정한다. 참조 화면에서 대응 블록(corresponding block)의 예측 모드(prediction mode)가 움직임 보상 모드로 일치하면, 현재 블록 또한 동일한 모드로 예측된다. 그 후, RD-비용(RD-cost)을 주어

진 임계값(Threshold)과 비교하여 이를 만족시키지 못하면 다른 방향으로 예측 과정을 다시 수행한다.

Ding 등은[5] 이웃하는 블록과 참조 화면의 대응 블록에서 SKIP 모드와 Intra 모드 및 시점 간 예측 모드의 RD-비용을 비교하여, 조건이 맞을 경우 움직임 추정 및 보상 과정을 생략함으로써 연산량을 감소시키는 기법을 제안하였다. 또한 시점 간 예측 과정에서 얻은 움직임 벡터(motion vector)를 재활용하여 현재 블록의 움직임 벡터로 활용하는 방법도 제시하였다.

실제 카메라의 위치와 주어진 외부 변수(extrinsic parameter)를 이용하여 시점 간 예측 모드의 탐색 범위를 적용적으로 조절하는 기법들도 연구되었다. Lu 등이 [6] 제안한 기법은 실제 카메라의 정보를 기반으로 탐색 창의 크기와 모양을 제어하였다. 시점 참조 화면(view directional reference frame)에서 현재 블록에 대응하는 블록은 에피폴라선(epipolar line) 상에 위치하므로 이를 이용하여 탐색 범위를 조절하면 적은 연산으로도 정확한 예측을 할 수 있다. 이와 비슷한 개념으로 Kim 등은 [7] 예측한 변위 벡터(disparity vector)들의 임계값을 비교하여 적은 탐색 범위로도 효율적인 예측을 수행하는 기법을 제시하고 있다.

3. 객체 분할 기법을 통한 빠른 예측 모드 선택 방법

3.1 현 다시점 부호화에서의 예측 모드 선택 특성

다시점 영상 부호화에서는 화면 내 정적(stationary)이며 균일(homogeneous)한 영역에 대해 움직임 보상 예측을 적용함으로써 보다 나은 성능을 얻을 수 있다. 이에 반해 시점 간 예측은 화면 내 동적인 영역에 적합하다. 빠르게 움직이며 전경에 위치한 객체가 포함된 영역에 움직임 보상 예측을 적용할 경우, 8×8 이하 크기의 블록 모드(sub-block mode)로 예측되는 경우가 많다. 이와 같이, 움직임 보상 예측을 이용하여 전경의 객체를 예측하기 위해서는 다수의 움직임 벡터가 필요하므로 성능이 저하된다. 따라서 이러한 영역에 대해서는 시점 간 예측을 적용하는 것이 적합하다.

반면에 변위 벡터는 화면 내 객체의 움직임과는 연관이 없다. 변위 벡터의 크기는 카메라의 상대적인 위치에 의해서만 결정된다. 따라서 시점 간 예측을 이용하여 움직이는 객체를 예측할 경우, 적은 수의 변위 벡터로도 효율적인 예측이 가능하다. 그림 1에서 오른쪽 화면의 회색 블록들은 시점 간 예측이 적용된 영역을 나타낸다.

그림 1의 왼쪽 화면과 오른쪽 화면은 동일한 Ballroom 영상의 일부이다. 왼쪽 화면에서 확인할 수 있듯이, 빠르게 움직이는 전경의 객체는 주로 시점 간 예측



그림 1 시점 간 예측이 적용된 매크로블록의 분포도

이 적용된다. 이러한 사실을 토대로 전경과 배경의 객체를 정확하게 구분해낼 경우, 현재 부호화 대상 블록의 예측 모드를 미리 결정할 수 있다.

3.2 움직임 기반 객체 분할 기법

본 논문은 각 매크로블록이 속한 객체를 효과적으로 구분하기 위해 움직임 벡터와 SAD(Sum of Absolute Difference)를 이용하는 객체 분할 기법을 제안한다[8]. 움직임 벡터는 시간 참조 화면의 대응 블록과 현재 블록의 거리 차이를 나타낸다. 객체의 움직임이 빠를수록 움직임 벡터의 크기가 증가하므로 적절한 임계값(threshold)을 설정함으로써 움직이는 객체를 구분해낼 수 있다. 그림 2에서 회색 블록들은 움직임 벡터의 크기가 임계값을 초과하는 영역을 나타낸다.



그림 2 움직임 벡터를 이용한 객체 분할 결과 화면

현재 움직임 탐색(motion estimation) 과정에서 구한 움직임 벡터는 현재 블록과 가장 유사한 대응 블록을 가리키는 대신, Lagrangian 비용 함수(cost function)를 이용하여 계산된 최적화된 영역을 가리킨다[9]. Lagrangian 비용 함수는 (1)과 같이 SAD와 움직임 벡터의 크기를 동시에 고려하여 최적화된 대응 블록을 계산한다.

$$J(X_i) = D(X_i, M) + \lambda \cdot R(X_i, M). \quad (1)$$

(1)에서 X_i 는 i 번째 매크로블록을 뜻하고 M 은 최적화된 움직임 벡터를 나타낸다. Lagrangian 비용 함수로 계산된 움직임 벡터는 해당 객체의 실제 움직임 정도를 정확히 나타내지 않는다. 따라서 객체의 정확한 움직임을 나타내기 위해서는 현재 블록과 가장 유사한 대응 블록을 가리키는 움직임 벡터를 구할 필요가 있다. 이를 위해 움직임 탐색 과정에서 움직임 벡터를 구하는 방식을 (2)와 같이 수정한다.

$$MV = \operatorname{argmin}_{MV} \left\{ \sum_{k \in B_r, k' \in B_r} |I_c(k) - I_r(k')| \mid B_r \in SW_r \right\} \quad (2)$$

(2)에서 I_c 와 I_r 은 현재 블록과 대응 블록의 휘도 성분(intensity)의 값을 나타내고, SW_r 는 참조 블록에 대한 탐색 원도를 가르킨다. 참조 화면에서 휘도 성분의 차이가 가장 작은 블록을 대응 블록으로 정해서 움직임 벡터를 구한다.

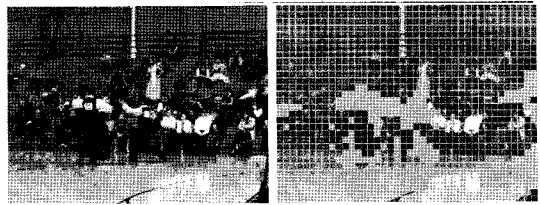


그림 3 동일 위치에서의 휘도 차이를 이용한 객체 분할 결과 화면

현재 블록과 동일한 좌표에서 측정한 SAD 값은 참조 화면 내 대응 블록과의 휘도 차이를 나타낸다. SAD 값을 이용할 경우 휘도 값이 균일한 영역에 크게 영향을 받지 않으므로 앞서 제시한 움직임 벡터 기반의 객체 분할 기법에서 발생한 오류를 보정할 수 있다. 그림 3은 SAD 값을 이용하여 객체를 분할한 결과를 나타낸다.

3.3 변위 기반 객체 분할 기법

변위 기반의 객체 분할 기법은 전경에 위치한 객체를 구분한다. 이러한 기법을 적용하여 움직임 기반의 객체 분할 기법을 적용했을 때 발생하는 오류를 보정할 수 있다. 변위 벡터는 객체의 깊이와 관계가 있으므로 변위 벡터의 크기가 커질수록 객체가 전경에 위치함을 알 수 있다. 따라서 변위 벡터의 크기에 임계값을 설정함으로써 전경의 객체를 효과적으로 구분할 수 있다.

변위 기반의 객체 분할 기법은 그림 4에서와 같이 계층적인 B 예측 구조(hierarchical-B prediction struc-

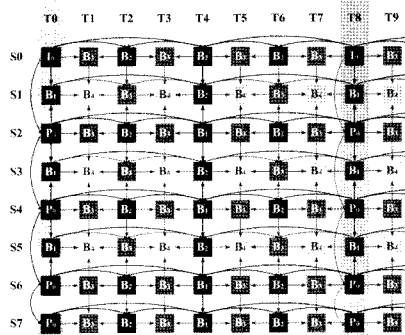


그림 4 계층적인 B 예측 구조

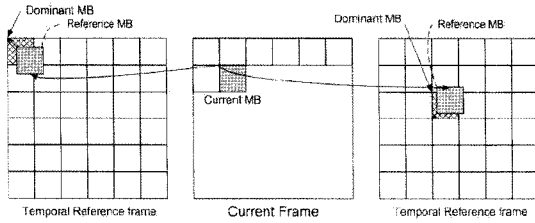


그림 5 시간 참조 화면의 우세 블록 탐색 과정

ture)의 앵커 픽처(anchor picture)에 대해 적용된다. 현재 부호화 대상 블록은 비앵커 픽처(non-anchor picture)에 위치하므로 앵커 픽처 상의 그리드(grid)에 대응하는 블록을 찾을 필요가 있다. 시간 참조 화면(temporal reference picture)에서 탐색한 대응 블록이 그리드 상에 위치하는 경우는 거의 없으므로, 우세 블록을 찾아 이를 대응 블록 대신 이용하는 과정이 필요하다. 이와 같이 앵커 픽처 상의 대응 블록을 탐색하고 나면 현재 블록이 속한 객체가 전경에 위치하는 지의 여부를 판단할 수 있다. 그림 5는 우세 블록을 이용하여 대응 블록을 대체하는 과정을 나타낸다.

만약 우세 블록과 현재 블록의 Intra 예측의 RD-비용 차이가 임계값을 넘을 경우, 우세 블록에 대한 잘못된 탐색이 이루어졌다고 보고 시간 참조 화면에 대응하는 블록이 없다고 정한다. 따라서 이전 움직임 기반의 객체 분할 기법의 결과만으로 객체를 구별하게 된다.

3.4 빠른 예측 모드 선택 과정

이 논문에서 제안한 움직임 기반 객체 분할 기법을 위해 움직임 탐색 과정이 비앵커 픽처에 대해 16×16 블록 크기로 수행된다. 이 과정에서 측정된 움직임 벡터와 SAD 값은 움직이는 객체를 구별하기 위해 사용된다. 변위 기반의 객체 분할 기법을 위해 변위 탐색(disparity estimation) 과정이 16×16 블록 크기로 수행되며 전경의 객체는 변위 벡터의 크기와 분포를 기반으로 구별한다.

이러한 객체 분할 기법의 결과를 바탕으로 현재 블록의 예측 모드를 미리 결정할 수 있다. 만일 현재 블록이 전경의 움직이는 객체에 속할 경우 시점 간 예측을 적용한다. 현재 블록이 배경에 속할 경우 움직임 보상 예측만을 이용하여 예측함으로써 부호화 시간을 효율적으로 줄일 수 있다.

4. 실험 결과

본 실험은 다시점 영상 부호화 표준의 참조 소프트웨어인 JMVM 6 [10]를 기반으로 1차원 평행 카메라 구조(1-D parallel)로 획득한 Ballroom과 Exit 영상에 대해 이루어졌다. 각 영상에 대해서는 다시점 영상 부호화

표 1 영상 정보

	특성	시점 수	카메라 배열
Ballroom	640×480, 25fps (Rectified)	8	1차원 평행 구조
Exit	640×480, 25fps (Rectified)	8	1차원 평행 구조

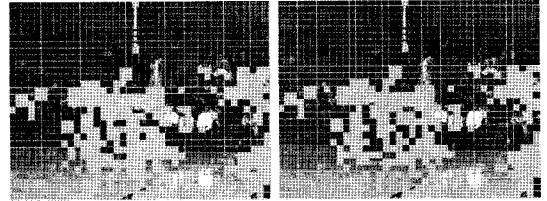


그림 6 Ballroom 영상에서의 시점 간 예측이 적용된 블록 분포(회색 블록은 시점 간 예측이 적용된 영역)

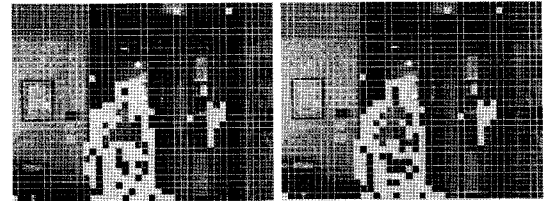


그림 7 Exit 영상에서의 시점 간 예측이 적용된 블록 분포(회색 블록은 시점 간 예측이 적용된 영역)

실험 조건에 맞춰 실험을 수행하였다[11]. 표 1은 각 영상의 특성과 실험 환경을 나타낸다.

그림 6과 7의 회색 블록은 Ballroom 영상과 Exit 영상에서 시점 간 예측이 적용된 영역을 나타낸다. 각 그림의 왼쪽 화면은 표준 참조 소프트웨어를 이용하여 실험을 수행한 결과를 나타낸다. 오른쪽 화면의 블록들은 이 논문에서 제안한 기법을 적용하였을 때 시점 간 예측이 적용된 영역을 나타낸다.

그림 6과 7에서 볼 수 있듯이, 본 논문에서 제안하는 기법을 적용한 결과는 표준 참조 소프트웨어의 결과와 거의 일치한다. 이러한 사실은 제안하는 기법이 화질의 큰 저하 없이 계산 시간을 효과적으로 감소시켰음을 보여준다.

표 2와 표 3은 각 Ballroom 영상과 Exit 영상의 PSNR (Peak Signal-to-Noise Ratio)과 비트율(bit-rate) 정보를 나타낸다.

표 4는 제안한 기법의 부호화 시간의 절감 비율을 나타낸다. 예측 과정을 대부분 생략함으로써 거의 절반에 가까운 부호화 시간을 감소시킬 수 있었다. 이는 객체 분할 기법에 소요된 시간을 더한 결과이다.

마지막으로 그림 8과 그림 9는 표 2와 표 3의 결과를

표 2 참조 소프트웨어에서의 결과

	Basis QP	PSNR (db)	Bit-rate (Kbit/s)
Ballroom	37	33.7	376.0
	32	35.3	583.1
	27	36.2	738.9
Exit	37	37.7	257.9
	32	38.3	325.9
	27	39.4	504.9

표 3 제안 기법의 결과

	Basis QP	PSNR (db)	Bit-rate (Kbit/s)
Ballroom	37	33.6	382.6
	32	35.3	589.4
	27	36.2	747.7
Exit	37	37.7	259.8
	32	38.3	326.8
	27	39.3	505.6

표 4 제안 기법의 시간 절감 비율

	평균 시간 절감 비율 (%)
Ballroom	-46
Exit	-43

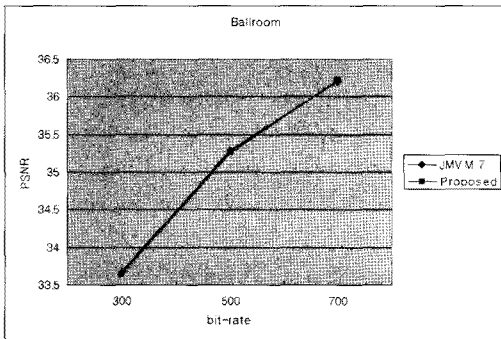


그림 8 Ballroom 영상에서의 PSNR 및 비트율 그래프

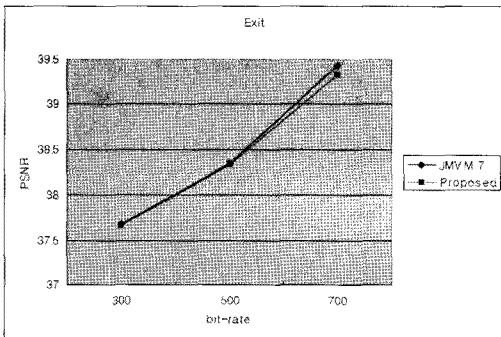


그림 9 Exit 영상에서의 PSNR 및 비트율 그래프

그래프로 나타낸다. 그림 8에서 알 수 있듯이 본 논문에서 제안한 기법은 Ballroom 영상에서 화질의 큰 손실 없이 부호화 시간을 단축했다. 그림 9의 경우 Exit 영상의 화질 저하는 변위 기반의 객체 분할 기법의 오류로 인해 일어난 것이다. 이는 영상 내 객체가 주로 전경에 위치하지 않기 때문에 발생한다.

5. 결론 및 향후 연구과제

본 논문은 화면 내 객체를 구분하여 이를 예측 모드 결정에 이용하는 기법을 제안한다. 빠르게 움직이는 전경의 객체는 시점 간 예측이 주로 적용된다는 사실을 토대로 해당 영역에 대해 예측 모드를 제한함으로써 다시점 영상 부호화 과정의 연산량을 효율적으로 감소시켰다.

이러한 객체 분할 기법은 움직임 기반 분할 기법과 변위 기반 분할 기법으로 이루어지며, 각 기법은 움직임 벡터와 변위 벡터의 특성을 이용하였다. 실험 결과, 평균 45% 에 가까운 부호화 시간을 단축할 수 있었으며 화질에는 거의 손상을 미치지 않았다.

향후 연구과제로는 영상 내 분할된 영역의 변위가 움직임보다 과도하게 크거나 작을 때를 가정하여 적합한 예측 모드를 선택하는 기법에 대해 연구한다.

참고 문헌

- [1] A. Smolic, P. Kauff, "Interactive 3-D video representation and coding technologies," Proc. IEEE, Special Issue on Advances in Video Coding and Delivery, Vol.93, No.1, pp. 98-110, Jan. 2005.
- [2] 호요성, 오관정, "다시점 비디오 부호화 기술 동향", 전자공학회지, 제34권, 제8호, pp. 916-917, 2007.
- [3] ISO/IEC JTC1/SC29/WG11, MPEG2006/W7798, "Description of core experiments in MVC," Jan. 2006.
- [4] Xiaoming Li, Debin Zhao, Xiangyang Ji, Qiang Wang, Wen Gao, "A fast Inter frame prediction algorithm for multi-view video coding," IEEE International Conf. on Image Processing, Vol.3, 2007.
- [5] Li-Fu Ding, Pei-Kuei Tsung, Shao-Yu Chien, Wei-Yun Chen, Liang-Gee Chen, "Computation-free motion estimation with Inter-view mode decision for multi-view video coding," IEEE International Conf. on 3DTV Conference, pp. 1-4, May 2007.
- [6] Jiangbo Lu, Hua Cai, Jian-Guang Lou, Juang Li, "An epipolar geometry-based fast disparity estimation algorithm for multiview image and video coding," IEEE Transaction On. Circuits and Systems for Video Technology, Vol.17, No.6, Jun. 2007.
- [7] Yongtae Kim, Jiyoung Kim, Kwanghoon Shon, "Fast disparity and motion estimation for multi-view video coding," IEEE Transaction On. Consumer Electronics, Vol.53, No.2, May 2007.

- [8] 이서영, 신광무, 김기완, 정기동, “모드 생략 기법을 이용한 빠른 다시점 영상의 부호화 방법”, 한국정보과학회 한국컴퓨터종합학술대회 2008, Vol.35, No.01(A), pp. 237-238, Jun. 2008.
- [9] Sullivan, G.J., Wiegand, T., “Rate-Distortion Optimization for Video Compression,” IEEE Signal processing Magazine, Vol.15, issue 6, pp. 74-90, 1998.
- [10] ISO/IEC JTC1/SC29/WG11, JVT-Y207, “Joint Multiview Video Model (JMVM) 6.0,” Oct. 2007.
- [11] ISO/IEC JTC1/SC29/WG11, JVT-U211, “Common Test Conditions for Multiview Video Coding,” Oct. 2006.



이 서 영

2007년 부산대학교 전자전기정보컴퓨터공학부 졸업(학사). 2009년 부산대학교 컴퓨터공학과 졸업(석사). 2009년~현재 토필드 부설연구소. 관심분야는 영상 부호화 및 전송



신 광 무

2005년 부산대학교 전자전기정보컴퓨터공학부 졸업(학사). 2007년 부산대학교 컴퓨터공학과 졸업(석사). 2007년~현재 부산대학교 컴퓨터공학과 박사과정. 관심분야는 영상 부호화, 임베디드 리눅스



정 기 동

1973년 서울대학교 졸업(학사). 1975년 서울대학교 대학원 졸업(석사). 1986년 서울대학교 대학원 계산통계학과 졸업(이학박사). 1978년~현재 부산대학교 컴퓨터공학과 교수. 관심분야는 멀티미디어 시스템 및 통신, 병렬처리, 영상 부호화,

파일시스템