

스테레오스코픽 방송용 비디오 인코더 프레임구조 분석

정회원 권 동 현*, 이 태 흥*

Analysis of Video Encoder Frame Structure for Stereoscopic Broadcasting

Donghyun Kwon*, Taihong Yi* *Regular Members*

요 약

스테레오스코픽 비디오 영상은 좌측영상과 우측영상으로 이루어진 한 쌍의 영상으로 구성되며, 이들 좌우 영상들 사이에는 높은 상관관계를 가진다. 스테레오 영상의 압축을 위한 부호화는 이들 좌우 영상의 높은 상관관계와 각 시각 내에서의 이웃 픽처와의 시간적 상관 관계를 모두 고려하여 이루어진다. 이 때, 왼쪽 영상의 정보는 기저층에서 처리하고 오른쪽 영상 정보는 상위 층에서 처리하는 것이 일반적이다. 본 논문은 MPEG-2 다시점 프로파일을 이용하여 입체영상 부호화 코덱을 구현할 때, 상위 층의 픽처 처리순서에 따라 부호화 효율이 어떻게 달라지는지를 분석하여, 실제 시스템 운용을 위한 코덱 구조를 제안하였으며, DTV 및 HDTV급 스테레오 시퀀스를 사용한 실험을 통하여 검증하였다.

ABSTRACT

Stereoscopic images using one pairs of right and left eye images have high correlation between each other. Stereoscopic coding considers these high correlations between both images as well as the temporal correlations between neighbor pictures. In general, the information of the left eye is processed in base layer and that of the right eye is processed in enhancement layer. A structure for picture processing of enhancement layer in stereoscopic codec design using MPEG-2 multi-view profile is proposed in the sense of practical viewpoints. The coding efficiency of the proposed method is analyzed in accordance with picture processing order and the experiments are performed on DTV and HDTV stereoscopic sequence.

I. 서 론

텔레비전 방송 기술은, 흑백시대에서 컬러시대를 거친 것과 같이, 아날로그 시대에서 디지털 시대로 변화하고 있고, 그 중에서 특히 디지털 TV 방송 기술은 SDTV(standard definition TV)와 더불어 HDTV(high definition TV) 방송이 부분적으로 실시되고 있다. 현재 세계 각국의 디지털 방송 기술은 크게 미국, 유럽 일본 방식의 세가지 방식이 대세를 이루고 있으며, 기술적인 측면에서 비디오의 경우 MPEG-2 기술이 채택되었고, 오디오의 경우

MPEG-2 혹은 DOLBY AC-3 기술이 채택되어 있는 상황이다.

그러나, 디지털 방송은 우리의 시각을 통해 매일 보고 경험하는 3차원 자연 영상을 2차원 평면 화면에 디스플레이하기 때문에 현장감, 실제감 면에서는 많이 떨어지는 것이 사실이다. 이러한 부족한 면을 채워주기 위한 기술로 3D TV방송이 고선명TV 방송의 뒤를 이을 새로운 영상 매체로 등장할 것으로 예측되고 있고, 우리나라도 일본과 같이 2002년 월드컵 때 3D TV 시험 방송을 실시할 계획으로 현재 연구가 진행되고 있다.

* 영남대학교 대학원 전자공학과 영상통신연구실(kdh@lily.ee.yeungnam.ac.kr),
논문번호 : 010060-0410, 접수일자 : 2001년 4월 11일

3D TV는 비디오 전화, 가상현실, 로보틱스 등 많은 응용 분야에 적용될 수 있다. 3D 영상과 관련된 연구는 과거에 주로 3D 영화, 홀로그래피(holography), 사진기술 등에 관해 이루어졌고, TV 방송에 관한 연구는 거의 이루어지지 않았다. 그 후 고선명TV에 관한 기술이 발전하고 성숙되어 고선명 TV의 높은 해상도를 이용하는 3D TV 연구가 활성화되었다.^[1]

3D TV 연구는 크게 자연에 존재하는 입체적 세계를 어떻게 영상으로 표현하는가 하는 입체 카메라 부분과 인간의 시각이 3차원의 세계를 인지하도록 어떻게 디스플레이 하는가 하는 디스플레이 관련 부분 및 전송을 위해 얼마나 효율적으로 대역을 압축하는가 하는 부호화 관련 부분으로 나눌 수 있다.

그 중 디스플레이 관련연구는 특수안경을 끼고 일반 모니터를 통해 입체감을 느끼도록 하는 스테레오스코픽 디스플레이 부분과 특수 모니터를 사용해 안경 없이 입체감을 느낄 수 있게 하는 오토스테레오스코픽(auto-stereoscopic) 디스플레이 부분으로 나눌 수 있다. 현재의 기술에서 스테레오스코픽 디스플레이 기술이 화질면에서 우수하며 경제적 측면에서도 효율적 운용이 가능하여 많이 사용되나, 방송용으로는 적합하지 않아 특수 안경 없이 디스플레이 하는 기술의 연구도 활발히 진행 중인 상태이다. 따라서 가까운 미래에는 특수 안경이 필요 없이 디스플레이되는 오토스테레오스코픽 영상이 보편화될 것으로 기대되고있다.

디지털 TV 영상의 대역압축은 2차원 비디오 영상프레임 내에서의 공간적 상관성을 줄이는 방법과 시간 축 방향에서의 상관성을 감소시키는 방법을 동시에 사용하여 이루어질 수 있다. 현재 TV 영상은 2:1에서 수백 대 1의 비율로 압축되며 이것은 MPEG-2 표준안^[2]을 따르고 있다.

3D TV에 사용되는 스테레오 영상의 압축 방법은 MPEG 그룹에서 MPEG-2 MVP(multiview profile)^{[3][4]}로 규정하여 표준화 작업을 마무리하였다. MVP는 두 개의 계층 구조로, 기저 층(base layer)에서는 왼쪽 눈에 대한 부호화를 행하고, 상위 층(enhancement layer)에서는 오른쪽 눈에 대한 부호화를 행하고 있다.

MVP를 적용한 코덱에서 기저 층은 계층구조를 사용하지 않는 알고리즘과 동일한 방법으로 영상을 처리하는 반면 상위 층에서는 영상의 처리순서 및 영상의 종류 등을 기저 층을 처리할 때에는 달리하

여 처리하고 있다.

본 논문은 MPEG-2 MVP를 적용한 스테레오 영상 부호화기를 구현할 때 가장 효율적인 압축을 위한 상위 층 프레임 구조를 제안하였다. 이를 위하여 상위 층에서의 영상 처리 순서 및 처리하는 영상의 종류에 따른 부호화율을 조사하였고, 실험을 통하여 입체영상 코덱에 사용될 상위 층에서의 여러 영상 프레임 부호화 방법을 비교하였다.

2장에서는 본 논문에서 사용하는 방법인 MPEG-2 MVP를 기술하였고, 3장에서는 처리하는 영상의 종류 및 순서에 대해 본 논문에서 제안한 방법을 기술하였으며, 4장에서는 실험 결과를 기술하여 각 방법을 비교한 후, 마지막으로 결론을 도출하였다.

II. MPEG-2 MVP(multi-view profile)

MVP는 ITU-T Rec. H.262 | ISO/IEC 13818-2 내에서 다중시점(multiple viewpoints) 관련 응용에 적절한 프로파일(profile)로 고안한 것으로, 스테레오스코픽 영상을 지원한다. MVP에서 좌측 카메라를 통한 정보는 기저 층에, 우측 카메라를 통한 정보는 상위 층에 할당된다. 기저 층은 모노스코픽(monoscopic) 부호화와 같은 방식이 사용되며, 상위 층에서는 시간 계위를 이용하여 움직임 정보와 시각차(disparity) 정보의 복합에 의한 예측을 이용한다.

동시 방송(simulcast) 동의 영역에 적용되는 시간 계위(temporal scalability)는 기존의 부호화 방식을 사용하는 기저 층 외에 상위 층에서 시간 축 상의 해상도를 높여주기 위해 사용하는 방법으로 부호화하는 층간의 공간 해상도는 같게 유지한다.

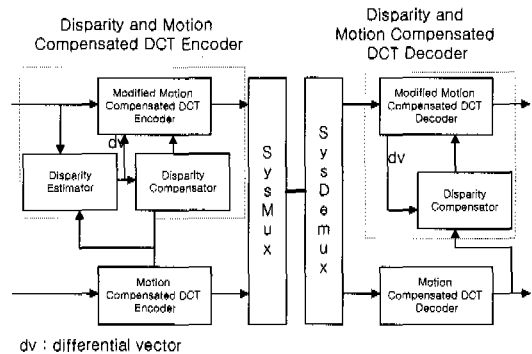


그림 1. 시간계위를 이용한 입체 영상 코덱 구조

그림 1은 시간 계위를 이용한 입체 영상 코덱의

구조를 나타낸 것이다. 기저 층인 왼쪽 카메라의 정보를 처리하는 부분은 계위를 사용하지 않고 부호화하는 방식과 동일한 방식을 이용하여 출력 비트 스트림을 생성한다. 반면 오른쪽 카메라에서 들어오는 정보의 처리를 담당하는 상위 층에서는 움직임 예측 및 보상 시 상위 층 정보에 기저 층의 정보를 함께 이용함으로써 부호화 효율을 극대화하여 상위 층의 비트 스트림을 생성한다.

MVP에서의 vbv 버퍼크기와 부호화율 등은 Main 프로파일과 차이가 난다. 먼저, MVP에서의 vbv (video buffering verifier) 버퍼 크기는 표 1의 버퍼 요구사항과 같이 각 레벨 기저층의 버퍼는 계위를 사용하지 않는 Main 프로파일에서의 버퍼와 같은 크기를 가지나, 상위 층에서는 1.5배 이상의 큰 버퍼를 가진다는 것을 알 수 있다.^[4]

표 1. VBV 버퍼 크기 요구사항

Level	Layer	Main	Multiview
High	Enhancement 1	9 781 248	15 898 480
	Base		9 787 248
High1440	Enhancement 1	7 340 032	12 222 464
	Base		7 340 032
Main	Enhancement 1	1 835 008	3 047 424
	Base		1 835 008
Low	Enhancement 1	475 136	950 272
	Base		475 136

표 2. 다중시점 프로파일@메인, 하이 레벨

Profile and level Indication	MP@ML	MVP@ML	MP@HL	MVP@HL
No. of Layers				
Layer id	0	1	0	1
Scalable Mode	Base	Temporal	Base	Tempora
Maximum sample density (H/V/F)	720/576/30	720/576/30	1920/1152/60	1920/1152/60
Maximum sample rate	10 368 000	10 368 000	62 668 800	62 668 800
Maximum total bit rate /1000000	15	25	80	130

표 2에서는 Main 레벨과 High 레벨에서 MVP의

최대 부호화 율 등을 나타내었다. DTV 급에서 기저 층은 15 Mbps가 최대이나 시간 계위를 사용하면 25 Mbps까지 허용된다는 것을 알 수 있고, HDTV 급에서는 기저 층에서 80 Mbp, 상위 층에서는 130 Mbps까지 허용된다는 것을 알 수 있다. 샘플링 율 등은 기저 층이나 상위 층 모두 호환성을 유지한다는 것을 알 수 있다.

그 외에도 MVP에서는 영상을 얻을 때 카메라에서 사용한 정보를 실을 수 있는 카메라 파라미터 확장부가 포함되어 카메라의 방향, 3차원 좌표에서의 위치, 초점 거리 및 렌즈 관련 변수 등을 전송할 수 있도록 하고 있다.

III. 3D TV 방송을 고려한 상위층 처리방법

입체영상을 구성하는 최소의 단위는 스테레오 영상이다. 스테레오 영상 압축을 위한 여러 방법이 존재할 수 있으나 가장 간단한 방법으로는 좌우 시각 정보를 완전히 독립적으로 보고 처리하는 것이다.^[5] 그림 2에서는 이를 나타내었는데 상위 층과 기저 층에서의 처리를 모두 독립적으로 하여 각 층이 비계층적 부호화 방법과 동일한 방법으로 처리하도록 하고 있다.

그림 3에서는 움직임 추정 시 기저 층의 처리 영상을 이용하여 압축 율을 높이려는 시도를 나타내고 있다. 즉 상위 층의 움직임 추정 시 기저 층의 정보를 이용해 부호화 비트 량을 줄인다. 그러나 그림 3과 같은 방법은 상위 층의 정보 처리를 위하여 기저 층만의 정보를 사용하기 때문에 기저 층과 상위 층의 정보를 모두 사용하는 방법에 비해서는 효율이 떨어진다.

따라서, 그림 4와 그림 5에서는 상위 층의 정보 처리를 위하여 기저 층의 정보 뿐만 아니라 상위 층의 정보까지도 동시에 이용하여 부호화를 하고 있다. 즉 같은 상위 층의 정보인 움직임 보상 정보에 기저 층과의 차이 정보인 시각차 정보를 동시에 이용하여 부호화를 시도하고 있다.

그림 4에서는 상위 층의 예측에 첫 프레임을 P(predicted) 영상으로 한 경우를 나타내고, 그림 5에서는 첫 프레임을 B(bi-directional predicted) 영상으로 처리하게 하였다.

또한 그림 4나 그림 5와 같이 처리를 하더라도 전송 과정에서 한 프레임의 정보에 에러가 생길 경우 그 뒤에 오는 프레임들의 정보에 영향을 주게 되어 연속적인 화질 저하의 우려가 크다는 문제가

있기 때문에 MPEG에서 계층 부호화를 수행하는 기본 틀에 맞게 GOP(group of pictures) 단위에서 차폐되게 처리해야 한다. 즉 GOP의 시작 프레임을 그림 4나 그림 5에서의 시작 프레임과 같이 이전 GOP 정보와는 독립적으로 처리하여야 한다.

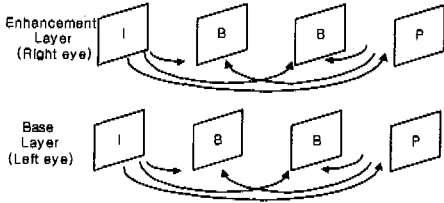


그림 2. 각 층의 독립적 부호화

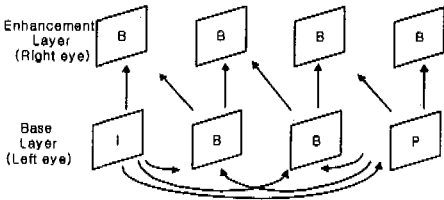


그림 3. 시각차 예측 정보만을 이용한 예측

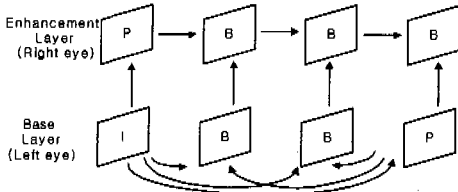


그림 4. 시각차와 움직임 보상을 이용한 예측(1)

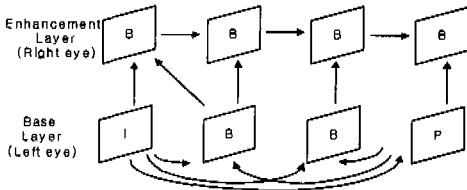


그림 5. 시각차와 움직임 보상을 이용한 예측(2)

IV. 실험 및 실험 결과

입체 영상 부호화에서 상위 층의 영상처리 순서 및 처리하는 영상의 종류에 따른 부호화 결과를 알아보기 위하여 그림 2와 같이 각 층을 독립적으로

부호화한 결과와 그림 4와 같이 P영상이 처음 처리되면서 시각차 정보 및 움직임 예측을 모두 사용하는 방법, 그림 5와 같이 B영상이 처음 처리되면서 움직임 보상 및 시각차 정보를 모두 이용하는 방법에 대해 각각 부호화를 행한 후 결과를 비교하였다.

부호화에 사용된 영상은 국제표준화기구에서 MVP 시험용 스테레오 영상으로 배포하고 있는 영상인 720×480 크기의 Street Organ 영상을 이용하였다.

부호화 시, GOP 크기는 N=15를 그리고 P 및 I 영상의 거리로는 M=3을 사용하였다. 실험을 위한 비트율은 비교를 위해 기저 층과 상위 층 모두에 대해 각각 4 Mbps의 율을 적용하였다.

표 3은 각 층에서의 vbv_delay를 비교한 것으로 처리하는 각 프레임에서 얻어진 vbv_delay를 프레임 수에 대해 평균한 값이다. 기저 층 및 독립적 부호화를 행하는 상위 층에 비해 시각차 정보와 움직임 보상정보를 모두 이용하는 상위 층의 버퍼 지연이 크다는 것을 알 수 있다.

표 4는 재생된 영상과 원 영상의 화질을 비교한 것으로 비교 대상이 되는 상위 층의 결과를 나타내었다. 독립적 부호화에 비해 기저 층의 시각차 정보를 이용하는 방법들이 대체적으로 우수한 결과를 나타내며, 처음 처리하는 영상이 P영상인 경우가 약간 더 나은 결과를 나타낸다는 것을 알 수 있다.

표 3. 평균 vbv_delay

Method	Base	Enhancement
Indep	34115	34006
B-first	34115	60077
P_first	34115	59026

표 4. 평균 PSNR(dB)

Method	Base	Enhancement
Indep	34.66	34.43
B-first	34.66	34.86
P_first	34.66	34.88

그림 6에서는 기저 층과 상위 층에서의 버퍼 상태를 각각의 방법에 대해 비교하였다. 중간 정보를 이용하지 않고 독립적으로 부호화를 행하는 경우 기저 층과 유사한 버퍼 상태를 나타내고 있으며, 층

간 정보를 이용하는 상위 층에서의 버퍼 상태는 기저 층의 버퍼 상태에 비해 높은 수준을 유지하지만 최대 허용 범위 내에서 적절하게 제어되고 있다는 것을 알 수 있다.

그림 7에서는 각 영상에 대한 PSNR을 프레임별로 비교하기 위하여 도표로 나타낸 것이다. 독립적인 부호화를 수행할 때 I영상인 경우 높은 PSNR을 나타내지만 P 및 B영상에서의 낮은 PSNR로 인하여 전체적인 평균 PSNR은 낮다는 것을 알 수 있으며, 처음 P영상이 처리되면서 시각차 정보를 함께 이용하는 경우가, 처음 B 영상이 처리되면서 시각차를 이용하는 경우보다 전반적으로 우수한 결과를 보인다는 것을 알 수 있다. 이는 P 영상에 할당된 비트가 B 영상에 할당된 비트보다 많아 결과적으로 코딩된 결과 화질에 영향을 주는 것으로 추측될 수 있다.

더불어 DTV 급 스테레오 비디오 영상이 아닌 HDTV 급 1920×1080 크기의 비디오 시퀀스 Tulip Garden을 이용하여 P 영상으로 시작하는 방법과 B 영상으로 시작하는 방법에 대해 그 결과를 비교하였다. HDTV급 스테레오 영상으로 처리하는 이유는 인간이 실재감(sensation of reality)을 잘 느끼는 화면 크기는 HDTV 크기이어야 하고, 깊이감(sensation of depth)을 잘 느끼기 위해서는 스테레오스코픽이어야 한다는 사실에 기인한다.^{[6][7]}

HDTV 스테레오 시퀀스를 이용한 실험은 N=13, M=3을 사용하여 좌우측 영상 모두에 18 Mbps를 적용하였다. 표 5의 결과를 볼 때 첫째 프레임은 B-프레임으로 처리한 것보다는 P-프레임으로 처리하는 것이 더 나은 결과를 보인다는 것을 알 수 있다. 그림 8의 프레임별 처리 결과를 보면 GOP의 처음 프레임에서 P의 결과가 더 나은 성능을 보여준다는 것을 알 수 있다.

이와 같이 GOP 단위로 GOP의 첫 프레임을 P 영상으로 처리하는 경우, B영상으로 처리하는 경우에 비해 DTV급이나 HDTV 급 모두 더 나은 결과를 나타낸다는 것을 알 수 있고, 전송 중 에러가 발생되더라도 GOP범위 이상으로 전파되지 않는다는 것을 알 수 있다.

표 5. 평균 PSNR

Method	Enhancement
B-first	32.5
P_first	32.9

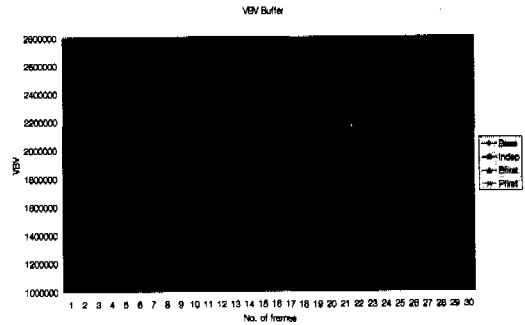


그림 6. vbv 버퍼 상태 비교도

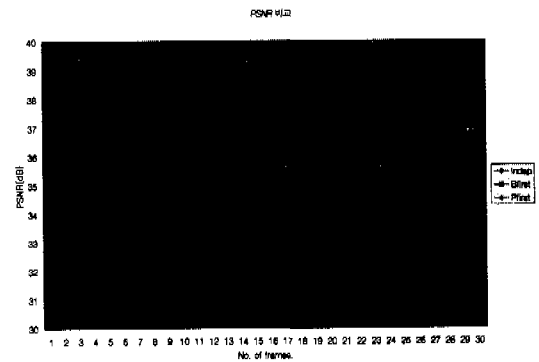


그림 7. 각 방법에 대한 PSNR 비교도

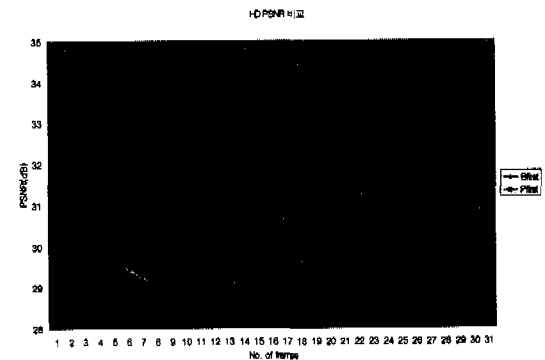


그림 8. HDTV 스테레오 영상의 PSNR 비교

V. 결론

3D TV 방송을 위한 스테레오스코픽 입체 영상 부호화 표준인 MPEG-2 MVP는 시간 계위 개념을 적용하여 좌측 시각 정보는 기저 층에서, 우측 시각 정보는 상위 층에서 처리한다. 본 논문에서는 MPEG-2 MVP를 이용하여 입체 영상 부호화 시스

램을 구현할 때 상위 층에서 처리하는 영상의 종류와 영상 처리 순서를 달리 했을 때의 부호화 결과를 DTV 및 HDTV급 시퀀스에 대해 분석하였으며 효율적이면서 전송에 적합한 형태의 프레임 처리 구조를 제안하였다.

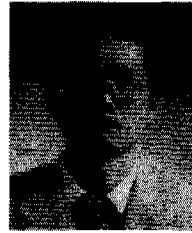
실험 결과, 시각차 정보 및 움직임 보상 정보를 동시에 사용하여 부호화를 행하는 경우가 좌우 시각을 독립적으로 부호화하는 경우보다 우수한 결과를 나타내었으며, GOP 내에서 P영상을 처음에 처리할 때와 B 영상을 처음으로 처리할 때의 결과를 비교하였을 때는 P영상을 먼저 처리하는 결과가 우위를 나타내는 것을 확인할 수 있었다.

참 고 문 헌

- [1] T.Motoki, H.Isono, I.Yuyama, "Present status of three-dimensional television research," *Proceedings of the IEEE*, vol. 83, no. 7, pp. 1009~1021, July 1995.
- [2] ISO/IEC 13818-2, "Information technology - Generic Coding of Moving Pictures and Associated Audio Information : Video," April 1996.
- [3] 권동현, 장현식, 김진웅, 이명호, 김영식, "MPEG-2 MVP를 이용한 입체영상 부호화기," *Proceedings of KITE Fall Conference 96*, vol. 19, No. 2, pp. 1191~1194, Nov. 1996.
- [4] ISO/IEC 13818-2, "Video Amendment 3," *ISO/IEC JTC1/SC29/WG11 N1366*, Chicago, Sep. 1996.
- [5] Do-Nyeon Kim and Myoung Ho Lee, "Compression of Stereoscopic Video By Using MPEG-2 Main Profile," *The 7th Int. Conf. on Signal Processing Appl. & Tech.*, vol. 1, pp. 117~121, Boston, Oct. 1996.
- [6] S. Yano and I. Yuyama, "Stereoscopic HDTV : Experimental System and Psychological Effects," *SMPTE Journal*, Jan. 1991.
- [7] N. Hiruma and T. Fukuda, "Accommodation Response to Binocular Stereoscopic TV Images and Their Viewing Conditions," *SMPTE Journal*, Dec. 1993.

권 동 현(Donghyun Kwon)

정희원



1990년 2월 : 영남대학교

전기공학과 졸업

1990년 8월~1992년8월 : 영남

대학교 전자공학과

석사 졸업

1992년 8월~1998년 6월 : 한국

전자통신연구원 연구원

1998년 9월~현재 : 영남대학교전자공학과 박사과정

<주관심 분야> 영상 및 비디오 압축/신호처리, 영상 /비디오 검색, HDTV 등

이 태 홍(Taihong Yi)

정희원

vol.25, no.1B 참조