

JSVC (Joint Scalable Video Coding) 표준화 동향

정세윤, 김규현 (한국전자통신연구원 디지털방송연구단)

I. 서론

본 고에서는 스케일러블 비디오 코딩 (Scalable Video Coding)과 관련하여 ISO/IEC와 ITU-T의 공동 비디오 부호화 표준화 그룹인 JVT (Joint Video Team)에서 진행되고 있는 JSVC (Joint Scalable Video Coding) 표준화 동향에 대해 소개한다.

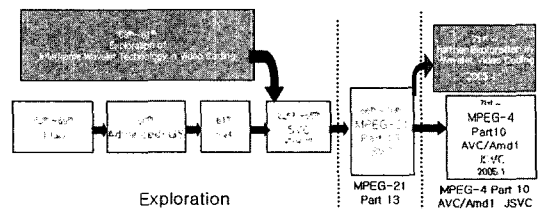
JSVC의 정식 표준 명칭은 ISO/IEC 14496-10 AVC:200x/AMD1|ITU-T H.264 Annex G 이다. JSVC 표준화 진행 상황은 2005년 1월에 WD 1.0이 (Working Draft) 제정되었으며 2005년 8월 현재 WD 3.0 까지 진행된 상태이며, 2006년 4월에는 CD (Committee Draft), 2006년 7월에는 FCD (Final Committee Draft), 2007년 1월에는 FDIS (Final Draft of International Standard)를 일정으로 진행 중에 있다.^[1]

SVC 표준은 이종 (heterogeneous) 네트워크 환경에서의 다양한 비디오 서비스를 위해 표준이 개발되고 있으며 화상 전화/회의, 모바일 비디오 스트리밍/방송, 디지털 비디오 감시 (Surveillance), 무선 홈 네트워크에서의

비디오 서비스 등 다양한 응용 서비스를 고려하고 있다.^[2]

본 고의 구성은 다음과 같다. 2절에서는 JSVC 표준화 이력에 대해서 설명하고 3절에서는 주요 기술 특징에 대해서 설명하고 4절에서는 현재 이슈화되고 있는 사항들에 대해 설명하고 끝으로 5절에서는 향후 전망에 대해서 설명하고자 한다.

II. SVC 표준화 이력



〈그림 1〉 SVC 표준화 이력

SVC 표준화 이력을 그림 1에 도시하였다. MPEG 표준화는 새로운 표준 기술에 대한 관련 기술 조사와 표준화의 필요성을 탐구 (Exploration)하는 과정을 거쳐서 정식 표준화가 시작된다. SVC 표준화의 출발은 탐구 단계에 있던 2 개의 그룹인, DCT 기반의 하

이브리드 부호화 (Hybrid Coding) 방식의 Advanced FGS 그룹과 Wavelet 기반의 Interframe Wavelet Video Coding 그룹의 2002년 10월 62차 MPEG 상하이 미팅에서의 SVC 그룹으로 연합된 것으로 부터라고 할 수 있다. SVC 그룹은 새로운 표준화를 위한 준비 작업을 진행하였고, 2003년 10월 MPEG 브리즈번 미팅부터는 MPEG-21 Part 13 SVC로 정식 표준화가 시작되었으며 Cfp (Call for Proposal)이 공고되었다. 2004년 3월 MPEG 뮌헨 미팅에서 제안된 기술 평가를 실시하였고 2004년 10월 MPEG 팔마 미팅까지 기술 경합이 진행되었으며 AVC|H.264를 기반으로 하는 독일 HHI 연구소의 제안 기술이 SVM 3.0 으로 채택되었다.

SVC는 MPEG에서만 표준화가 진행되고 있었으나 HHI의 제안기술이 AVC|H.264를 기반으로 한 기술이므로, AVC|H.264의 공동 표준화 제정 기구인 ITU-T에서도 AVC|H.264를 확장하는 SVC 표준화의 필요성을 인식하게 되어, 2005년 1월 MPEG 홍콩회의 부터는 SVC를 Video 그룹에서 JVT 그룹으로, MPEG-21 Part 13 SVC에서 JSVC로 명칭을 변경하여, 표준화가 두 기구의 표준으로 확대되어 공동으로 진행되고 있다.

2004년 3월 68차부터 2005년 7월 73차 MPEG 회의에서의 SVC 관련 주요 사항들을 표 1로 요약 하였다. 68차 MPEG 뮌헨 회의에서는 Cfp에 대한 제안 기술들에 대한 기술 평가가 최초로 진행된 회의로 AVC 기반 기술 진영과 Wavelet 기반 기술 진영이 경합을 벌였으며, 우열을 가리기 힘든 상황이었다. 차기 회의를 위해 Wavelet 진영에서 가장 좋은 성능을 보인 MSRA의 S/W와 AVC 기반 진영

<표 1> SVC 관련 MPEG 회의의 주요 사항

	회의 주요사항	결과문서	Reference SW
68차 Munich	Cfp 제안 기술 평가 -AVC와 Wavelet 결합	SVM 1.0	MSRA HH
69차 Redmond	AVC 기반 기술 우세	SVM 2.0	MSRA HH
70차 Palma	AVC 기반 기술로 확정	SVM 3.0	SVM 3.0
71차 Hong	JSVC로 명칭변경	JSVM 1.0 WD 1.0	JSVM 1.0
72차 Busan	JSVC정식 회의	JSVM 2.0 WD 2.0	JSVM 2.0
73차 Poznan	표준화 일정변경	JSVM 3.0 WD 3.0	JSVM 3.0

에서 가장 좋은 성능을 보인 HHI의 S/W가 Reference S/W로 공개 되었으며 SVM (Scalable Video Model) 1.0 이 제정되었다. 69차 MPEG 레드먼드 회의에서는 AVC 기반 기술 진영의 HHI 제안 기술이 가장 우수한 기술로 판명되었으나, 테스트 조건에 최적화한 실험 결과를 제출 하였기에 테스트 조건을 랜덤하게 지정할 수 있게 변경하여 차기 회의에서 다시 한번 경합을 하게 되었다. 70차 MPEG 팔마 회의에서도 AVC 진영의 HHI 제안 기술이 가장 우수한 기술로 판명 되었고 SVM 3.0의 기술로 채택되었으며, Wavelet 기반 진영 기술들은 Wavelet Coding 그룹으로 분리되었다. 70차 회의 이후부터는 기술 경합은 종료되고, SVM의 성능을 개선하는 것을 목표로 표준화가 진행되었으며, 71차 MPEG 홍콩 회의부터는 JSVC로 명칭이 변경되었고, AVC|H.264 표준 확장하는 방향으로 변경되었어, AVC|H.264의 기존 기술들과 융합도 고려하여 표준화가 진행되고 있다.

III. JSVC 주요 기술

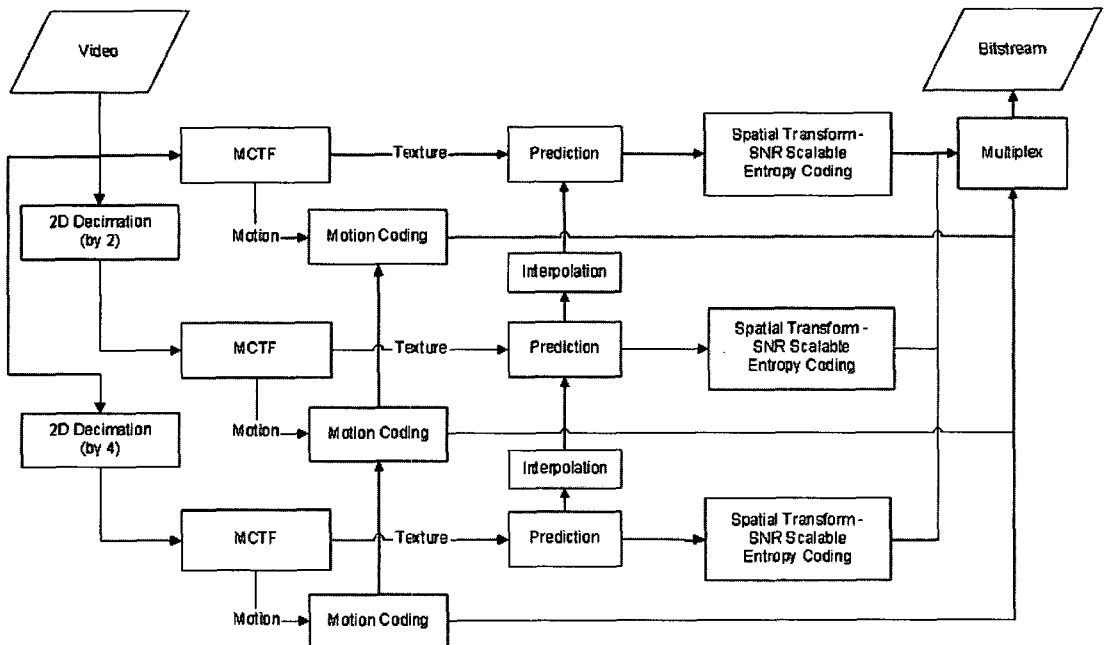
그림 2는 3 단계(level) 공간 스케일러빌리티를 제공하는 경우의 JSVC 인코더 구조도이다. JSVC의 특징은 공간 스케일러빌리티는 계층(Layer) 구조를 이용해서 제공한다는 것이다. 그림 2에서와 같이 각 공간 해상도 별로 계층을 구성하고, 각 계층 별로 부호화 처리를 하는것을 기본으로 하고 있다. 또 다른 특징은 MCTF (Motion Compensated Temporal Filtering)을 사용하여 시간적(Temporal) 스케일러빌리티를 제공한다는 것이다. 기본 계층(Base Layer)은 H.264와 호환성을 유지해야 하므로 Hierarchical B pictures를 MCTF 대신 사용한다. SNR 스케일러빌리티는 2 종류가 제공 가능하며, 임베디드(Embedded) 양자화(Quantization) 기법

을 이용한 CGS (Coarse Grain Scalability)와 비트플레인 (bitplane) 산술 (Arithmetic) 부호화 기법을 이용한 FGS (Fine Grain Scalability)가 제공 된다.

SVC는 공간, 시간, SNR 스케일러빌리티를 동시에 제공하는것을 기본으로 하나 한가지 스케일러빌리티만 제공하는 방식으로도 사용이 가능하다. 이하 세부 절에서는 JSVC의 공간, 시간, SNR 스케일러빌리티와 관련된 기술에 대해 상세히 살펴보고자 한다.

1. 계층 구조와 공간 스케일러빌리티

공간 스케일러빌리티는 MPEG-2에서 사용되었던 계층 구조를 통해서 제공이 된다. SVC에서는 다수의 계층을 사용이 가능하고 각 계층간의 해상도 차이가 2배가 아닌 경우



〈그림 2〉 SVC 인코더 구조

(non-dyadic) 도 지원하는 것이 특징이다. 각 계층은 입력 영상을 각 계층의 공간 해상도에 맞게 다운 샘플 필터링을 적용하여 입력을 받게 된다. 2의 배수가 아닌 경우에는 다운 샘플 필터링에서 크로핑 (cropping) 까지 포함 한다.

JSVC는 AVC1H.264에 계층 구조가 추가된 것으로 볼 수 있으므로, AVC1H.264의 프레임내 예측 (intra prediction)과 프레임간 예측 (inter-frame prediction) 2가 방법에 추가로 계층간 예측 (inter-layer prediction)이 추가된 것으로 봐도 무방하다. 계층간 예측은 움직임 벡터, 텍스처(texture) 데이터, 여분데이터 (residual) 에 대하여 사용하고 있다. 계층간 예측에서는 상위 계층의 매크로 블록 (Macro Block) 에 대응하는 하위 계층의 매크로 블록의 부호화 방식에 기반하여 예측이 적용된다.

하위 계층의 대응 매크로 블록이 프레임간 예측으로 부호화 된 경우, 상위 계층의 해당 매크로 블록더 프레임간 예측방식으로 부호화가 되며 움직임 벡터는 하위 계층의 움직임 벡터를 스케일링하여 사용한다. 이때 스케일링된 움직임 벡터를 그대로 사용하는 방식을 'base layer mode' 라고 하고 스케일링된 벡터로부터 1/4 화소 영역 내에서 추가로 탐색하여 사용하는 방식을 'quarter sample refinement mode' 라고 한다.

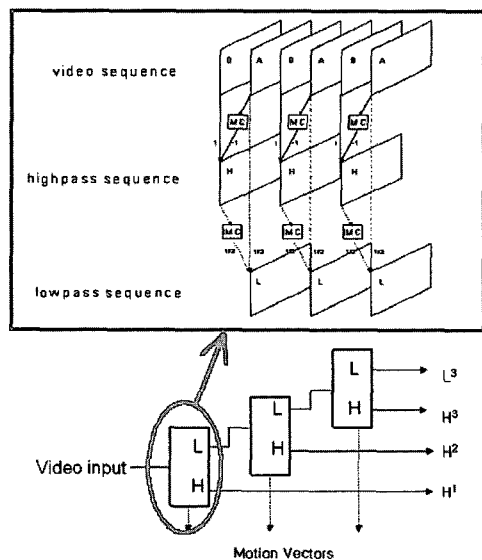
하위 계층의 대응 매크로 블록이 프레임내 예측으로 부호화 된 경우, 상위 계층의 해당 매크로 블록은 프레임내 예측으로 부호화 하는것을 기본으로 하며, 이때 하위 계층의 텍스처 데이터를 스케일링 하여 예측에 사용한다. JSVC 레퍼런스 S/W 에서는 하위 계층의

대응 매크로 블록의 부호화 방식에 상관없이 계층간 텍스처 예측을 사용 하는 것도 사용 가능 하며, 하위 계층의 대응 매크로 블록이 프레임내 예측이고 로우 패스 프레임(Key 프레임 이라고도 함)인 경우에만 사용하는 방식도 제공한다. 디코더의 메모리 크기를 고려하여 적절한 방법을 선택하여 사용할 수 있다.

계층간 여분 데이터 예측은 계층간에 움직임 예측이 사용된 경우에 추가적으로 적용이 가능하며, 두 공간 계층간의 대응 매크로 블록의 여분 데이터의 상관성도 높기 때문에, 여분 데이터에 대해서도 예측을 하면 압축 효율을 더욱 높일 수 있기 때문에 사용된다.

2. MCTF와 Hierarchical B Pictures

JSVC에서 시간 스케일러빌리티는 MCTF를 이용해서 제공된다. 그림 3은 MCTF의 개념

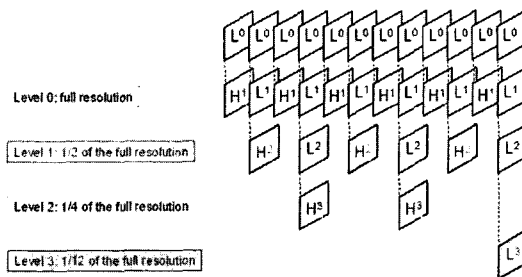


(그림 3) MCTF 개념도

도이다. JSVC에서 MCTF는 리프팅 (lifting) 구조를 사용하여 처리하고 있다. 리프팅 구조에서는 입력 영상 A와 B영상에 먼저 하이패스 필터링을 하여 하이패스 영상들을 구하며, 이 과정을 프리딕션 (prediction) 이라고 한다. 다음 하이 패스 영상과 B 영상에 로우패스 필터링을 하여 로우패스 영상을 구하며, 이 과정을 업데이트 (update) 라고 한다. MCTF를 한번 적용한 결과를 1차 시간 레벨 (temporal level)이라고 하며 H^1 , L^1 과 같이 위첨자로 시간 레벨을 표시한다.

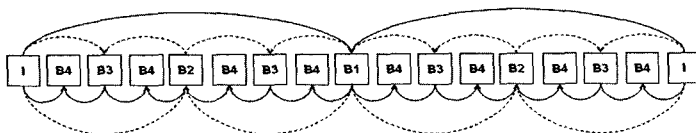
L 영상들에 대해서만 MCTF가 계속 적용되며, L 영상이 한 장이 될 때까지 MCTF가 반복 적용된다. Haar MCTF의 경우 L 영상의 개수는 1/2씩 줄어든다.

MCTF에서 시간 스케일러빌리티가 어떻게 제공되는지 그림 4에 도시되어 있다.

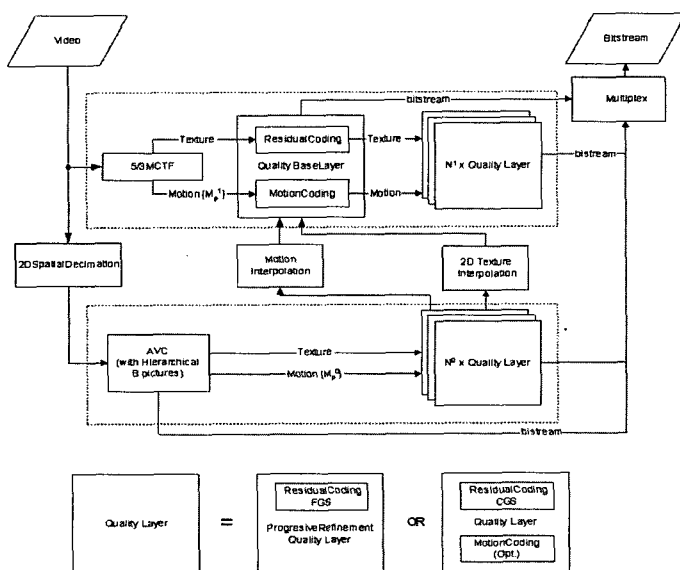


〈그림 4〉 MCTF와 시간 스케일러빌리티

시간 레벨이 그림 4에서는 0부터 4까지로 구성되어 있다. 모든 시간 레벨 데이터를 사용될 경우 전체 시간 해상도가 제공이 되며, 레벨 1부터 3까지 데이터를 사용할 경우 1/2



〈그림 5〉 Hierarchical B picture 구조



〈그림 6〉 JSVC의 CGS와 FGS

시간 해상도가 레벨 2부터 3까지의 데이터만 사용할 경우에는 1/4 시간 해상도가 제공되며 레벨 3만 사용할 경우 1/12 해상도가 제공이 가능하다. MCTF에서는 시간 레벨 별 시간 해상도가 2의 배수가 아닌경우도 제공할 수 있다.

JSVC의 기본 계층은 AVC|H.264와 호환되는것을 원칙으로 한다. 그러므로, 기본 계층에서는 MCTF대신 Hierarchical B pictures 구조를 사용한다. Hierarchical B pictures 구조는 MCTF에서 업데이트 과정이 생략된 것과 이해할 수 있으며, JSVC에서도 이를 이용하여 구현하고 있다.

3. CGS 와 FGS

JSVC에서 SNR 스케일러빌리티는 CGS와 FGS 2 종류가 제공된다. SNR 스케일러빌리티는 그림 6에서와 같이 Quality 계층에 의해 제공되며 CGS 방식 또는 FGS 방식 중 하나를 사용하여 처리 된다.

CGS는 임베디드 양자화 기법을 이용해 구현된다. 이는 Quality 계층은 임베디드 구조로 구성이 되며, 최소의 Quality를 갖는 기본 계층을 먼저 부호화 하고 양자화 스텝 크기를 줄여가며 이전 계층과의 차이 데이터를 부호화 하는 방식으로 처리 된다.

FGS는 점진적인 (progressive) 리파인먼트(refinement) 방법을 적용하여 처리한다. CGS와 비슷하게 Quality 기본 계층을 먼저 부호화 하고 양자화 스텝 크기를 줄여가며 차이 데이터를 비트플레인 부호화 방식과 유사한 변형된 CABAC 부호화를 적용하여 처리하며 CGS보다 미세한 SNR 스케일러빌리

티가 제공 가능하다.

IV. 주요 쟁점 사항

JSVC의 주요 표준화 쟁점 사항으로는 MCTF에서의 업데이트 과정의 생략과 CAVLC의 지원 여부가 있다.

JSVC의 MCTF는 Open-loop 구조를 사용하고 있다. 이는 MCTF의 움직임 보상(compensation) 과정에서 기존의 Hybrid 방식과는 달리 재구성된(reconstruction) 영상을 사용하는 것이 아니라 원 영상을 사용하고 있기 때문이다. 이는 인코더와 디코더의 움직임 보상 과정에서 불일치가 발생한다는 것을 의미한다. MCTF의 업데이트 과정에서 이러한 불일치의 보상이 이루어 지고, MCTF의 반복 적용으로 인한 압축 이득이 크므로, 불일치로 인한 손실은 무시되고 있었다. 표준화가 진행되면서 압축 이득을 높이는 기술들이 채택이 됨으로 해서 MCTF의 업데이트를 생략해도 충분히 압축 성능을 유지할 수 있다는 주장이 73차 MPEG 포즈난 회의부터 JVT 그룹내에서 제기되고 있다. 이들은 MCTF의 업데이트는 압축 효율을 높이는 것에 비해 계산 복잡도와 디코더의 메모리 요구량이 지나치게 높다고 주장하고 있다.

JSVC에서 CAVLC도 사용하자는 기고가 73차 MPEG 포즈난 회의부터 제기되고 있다. 이는 JSVC의 기반이 되는 HHI에서 CABAC을 개발한 기관이기 때문에 현재까지는 CABAC만 JSVC에서 고려되었다. CABAC이 CAVLC보다 압축 효율은 높으나 계산 복잡도가 높고, 특히 하드웨어로 구현하는데 매우 어렵다는 단점이 있다. JSVC에서 CAVLC 지

원 여부에 대해서는 74차 MPEG 니스 회의에서 논의가 될 예정이다.

그 외 쟁점 사항으로는 저장 표준인 AVC 파일 포맷에 대한 JSVC 확장과 현재까지 압축 효율을 우선으로 하여 표준화가 진행되었으므로 에러 강인성(resilience)에 대한 기술과 비트율 제어와 양자화 스텝 결정 방법 등이 있다.

V. 향후 전망

본 고에서는 JSVC의 표준화 이력 및 주요 기술인 MCTF와 Hierarchical B Pictures 구조 및 CGS와 FGS에 대해 설명하였고 주요 쟁점 사항으로 MCTF에서의 업데이트의 생략과 CAVLC의 지원에 대해 고려가 되고 있다고 설명하였다. 본 필자는 MCTF의 업데이트는 생략이 유력하다고 예상하며 CAVLC는 사용이 되며 AVC/H.264의 경우와 비슷하게 별도의 프로파일로 CABAC과 CAVLC의 사용이 분리될 것이라고 예상한다.

JSVC는 새로운 표준으로 출발하였으나 표준화 진행과정에서 AVC/H.264의 후속 표준화로 변경되어 진행되고 있으며, 이는 기존의 AVC 표준에 계층 구조가 추가된 형태이므로 기존에 없던 계층간 예측과 관련된 기술들만 새로이 추가되는 형태로 표준화가 진행되고 있음을 의미하며 표준화도 매우 빠르게 진행되고 있다.

SVC 표준화는 유니쿼터스(ubiquitous) 네트워크에서의 사용을 목표로 하는 비디오 표준으로 표준화가 시작되었으나, 현재의 AVC 후속 표준화로 변경된 JSVC로는 이를 달성하기 어렵다고 생각하며, JSVC 표준화 이후에

도 SVC와 관련된 새로운 표준화가 계속 진행될 것이라고 전망한다.

참고 문헌

- [1] Jens-Raniner Ohm, "Introduction to SVC Extention of Advanced Video Coding", ISO/IEC JTC1/SC29/WG11 N7315, Jul. 2005.
- [2] MPEG requirements group, "Applications and Requirements for Scalable Video Coding", ISO/IEC JTC1/SC29/WG11 N6880, Jan. 2005.
- [3] JVT group, "Joint Scalable Video Model (JSVM) 3.0 Reference Encoding Algorithm Description", ISO/IEC JTC1/SC29/WG11 N7311, Jul. 2005.
- [4] JVT group, "Working Draft 3 of ISO/IEC 14496-10:2005/AMD1 Scalable Video Coding", ISO/IEC JTC1/SC29/WG11 N7311, Jul. 2005.
- [5] Leonardo Chiariglione, "Resolutions of 73rd meeting", ISO/IEC JTC1/SC29/WG11 N7265, Jul. 2005.
- [6] Heiko Schwarz, D. Marpe, T. Wiegand, "Comparison of MCTF and closed-loop hierarchical B pictures", JVT-P059, Jul. 2005.
- [7] Woo-Jin Han, "CE6 response of Samsung Electronics: in-depth comparison of closed-loop and open-loop MCTF structures", JVT-P084, Jul. 2005.

저자소개



정세운

1995년 인하대학교 전자공학과 학사
 1997년 인하대학교 전자공학과 석사
 1996년 ~ 현재 한국전자통신연구원 디지털방송연구
 단 선임연구원
 주관심 분야 스케일러블비디오코딩, 멀티미디어통신,
 디지털 대화형 방송



김규현

1997년 ~ 현재 한국전자통신연구원 대화형미디어연
 구팀 팀장
 2001년 ~ 2005年 MPEG 한국대표 단장 (HoD)
 2004년 ~ 현재 APT(Asia Pacific Telecommunity)
 TG3 의장
 주관심 분야 영상처리, 멀티미디어통신, 디지털 대화형
 방송