

H.323을 이용한 다자간 영상 채팅 시스템 개발

박한엽* · 나인호* · 최연성*

*군산대학교 전자정보공학부

목 차

- I. 서론
- II. 영상회의 또는 채팅 시스템 개발 동향
- III. 영상압축 H.263
- IV. 음성 부호화 표준 G.711,G.722,G.723
- V. 구현 및 실험
- VI. 결 론

I. 서론

멀티미디어 데이팅 시스템은 지역적으로 떨어진 2지점 이상의 다 지점간에 영상, 음성 및 데이터 정보를 효율적으로 송수신 할 수 있도록 하여 회의의 생산성 향상을 지원하는 시스템이다.

기본적인 방향은 ITU의 H.323 권고안[1]을 이용하여 문자뿐만 아니라 오디오, 비디오를 송수신하는 멀티미디어 데이팅 서비스 플랫폼을 개발하는 것이다.

ITU에서는 이러한 요구에 맞추어 1990년도부터 지금까지 여러 표준안을 발표하게 되었다. 1995년 실시간 데이터 컨퍼런싱에 대한 T.120 표준을, LAN 상에서 비디오 통신(Audiovisual)에 대한 H.323 표준을, POTS 모델 연결을 통해 고품질의 비디오와 오디오 압축을 위한 H.324 표준을, 1990년 ISDN 비디오 컨퍼런싱에 대한 H.320 표준을 개발하여 인준했다. 이러한 권고안을 따라 외국에서는 디지털 방송, 인터넷 폰, 다자간 채팅을 이미 상용화시킨 상태이다.

다자간 데이팅 시스템은 초고속 정보 전송을 위한 통신망, 멀티미디어 장비, 영상 소프트웨어 및 컴퓨터 기반 공동작업 지원 응용의 4요소가 결합될 때 사용자에게 보다 나은 서비스를 보장해 줄 수 있다. ITU에서 배포되는 권고에는 호 제어 방식(call signaling), 멀티미디어 데이터 포맷 정의, 공유 정보 처리 기술 등

이 포함되어 있으며, 이종의 영상 회의 시스템간의 상호 운용성(inter-operability)에 대한 표준을 제공하고 있다. 본 논문에서는 이러한 표준화된 기술을 이용하고 더 나아가 발전시켜 보다 안정된 플랫폼을 구현하는 것을 그 목적으로 했다. 본 논문에서는 H.320보다 범용성이 큰 H.323을 모델로 하며, 비디오 코덱부에서 실시간 MPEG-2와 H.263을 선택적으로 이용하여 보다 빠르고 고품질로 데이팅이 가능하도록 했다. 또한 서버의 과부하나 네트워크 트래픽을 고려하여 보다 안정적인 멀티미디어 데이팅 서비스를 실현하고자 했다.

II. 영상회의 또는 채팅시스템 개발 동향

1 다중 다자간 영상채팅 현황

영상채팅은 기존의 문자 위주의 채팅에서 벗어나 멀티미디어 시대에 맞춘 영상과 음성을 지원할 뿐만 아니라, 파일을 전송할 수도 있는 멀티미디어 기반의 채팅 시스템이다. 영상 채팅은 씨앤조이(www.seenjoy.co.kr)를 시작으로 전파되었는데 PC카메라의 보급으로 신세대들의 새로운 문화로 자리잡았다. 영상 채팅을 기존 문자 채팅에 멀티 미디어의 재미를 더하여 보다 다양한 신세대의 욕구를 만족시킨다.

최근에는 오 마이 러브(www.ohmylove.co.kr), 웹114(www.web114.co.kr), 하늘사랑(www.skylove.co.kr), 버디버디(www.budybudy.co.

kr), 챗러브 (www.chatlove.co.kr)등 주요 채팅 서비스 업체들이 최근 전국적으로 PC방이 확산되는 등 인터넷을 이용할 수 있는 인프라가 개선됨에 따라서 영상과 음성을 결합한 멀티미디어 채팅서비스를 잇따라 선보이며 네티즌을 끌어 모으고 있다.

2 H.323의 개요

권고안 H.323은 보장성 서비스품질(QoS)을 제공하지 않는 LAN상에서의 멀티미디어 통신을 위한 터미널, 장치, 그리고 서비스를 기술하고 있다. H.323은 Ethernet과 같은 패킷 교환 네트워크상의 멀티미디어 통신에 적용되고, TCP/IP, IXP/SPX 등에서도 운영된다. 또한 회의 입장, 멀티포인트 통신, 다른 형태의 네트워크상의 터미널 호환성을 제공한다.

H.323은 다양한 어플리케이션을 가지고, LAN, 기업 네트워크, MAN, WAN, dial-up에 의한 LAN과 인터넷 등 다양한 형태를 연결할 수 있다. 가장 흥미 있는 어플리케이션은 인터넷상에서의 화상회의이다.

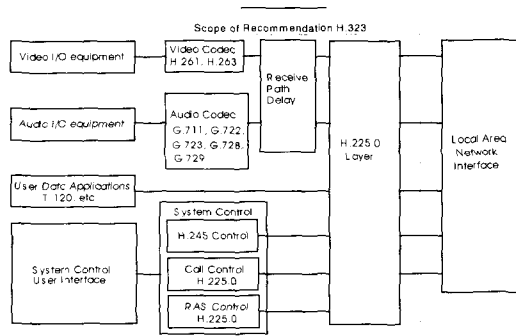


그림 1 H.323 시스템 구성도

H.323 터미널은 그림 1에서와 같이 사용자 인터페이스 장비, H.261, H.263 비디오 코덱, G.711, G.722, G.728, G.729, G.723 오디오 코덱, H.225.0 Layer, 시스템 제어기와 LAN 인터페이스 그리고 멀티미디어 통신 프로토콜 T.120시리즈[2]를 포함한다.

H.225.0은 호 신호, 회의 등록, 입장뿐만 아니라, 미디어 스트림 패킷화와 메시지 동기화를 H.245는 메시지들과 오디오, 비디오 그리고 데이터의 용량교환, 미디어 스트림을 위한 논리 채널 사용 절차를 제공한다.

H.323 터미널은 멀티포인트 구성에 사용될 수 있고, B-ISDN상에서 H.310 터미널, N-ISDN 상에서 H.320 터미널, GSTN상에서 H.324 터미널, ATM상에서 운영되는 H.310 터

미널 그리고 GSTN상의 V.70 터미널과 상호 작업할 수 있다. 그림 2는 H.323의 구성을 보여준다.

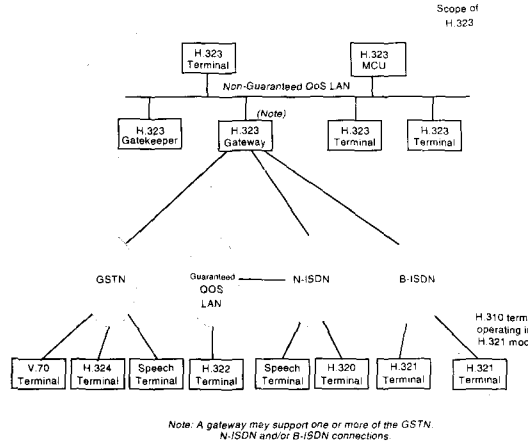


그림 2 H.323의 연결구성도

(1) H.323의 구조

H.323은 패킷교환망을 통한 음향 및 영상통신 서비스에 대한 기술적 요구사항을 다루고 있으며, 데이터 회의에 대한 T.120 규격을 참고하여 데이터 능력을 포함한 회의를 가능하게 한다. H.323의 범위는 망 자체나 다양한 망을 접속하는데 사용될 수 있는 전송 계층을 포함하지는 않는다. H.323은 망-기반 통신시스템을 위해 네 개의 주요 구성요소를 규정하고 있다. 단말(terminal), 게이트웨이(gateway), 게이트키퍼(gatekeeper), 그리고 MCU이다.

가. 단말(terminal)

단말은 고객측 종단 장비로 실시간 단방향 또는 양방향 통신을 제공한다. 모든 단말은 음성통신을 지원해야 한다. 영상과 데이터는 선택이다. H.323은 다른 음향, 영상, 데이터 단말이 같이 작동하는데 필요한 작동 모드를 규정하고 있어, 인터넷 전화, 음향회의 단말, 영상회의의 기술에 대한 차세대의 주도적 표준이 될 전망이다.

모든 H.323 단말은 채널 사용과 성능을 중재하는데 이용되는 H.245를 지원해야 한다. 그리고 세가지 부가적 구성요소는 호 시그널링과 호 설정을 위한 Q.931 축소 버전, 게이트키퍼와의 통신에 이용되는 프로토콜인 Registration/Admission /Status(RAS), 음향과 영상 패킷의 순서화를 위한 RTP/RTCP (Real-Time Protocol/Real-Time Control Protocol) 지원 등이다.

나. 게이트웨이(Gateway)

H.323 회의에서 게이트웨이는 선택 사항으로 H.323 회의 단말과 다른 단말간 변환 기능을 포함해 다양한 서비스를 제공하는데, 이 기능에는 전송 포맷(즉, H.225에서 H.221)간, 통신 절차(즉, H.245에서 H.242)간 변환이 포함된다. 또한, 게이트웨이는 음향과 영상 코덱 사이를 전환하고, LAN과 WAN측에서 호 설정과 해제를 수행한다. 그림 3은 H.323/H.320 게이트웨이를 나타내는데, H.323/H.320 게이트웨이가 있을 때, 데스크톱 H.323 시스템을 보유한 이용자는 전세계의 다른 오래된 H.320 시스템 이용자와 링크 될 수 있다.

다른 게이트웨이 유형도 역시 가능한데, 몇몇 ViDe 사이트는 H.323과 아날로그 사이의 게이트웨이와, H.323 단말 스테이션의 VGA 디스플레이를 취하고 방송 영상 망으로 보내기 위해 그것을 아날로그나 디지털 영상으로 전환하는 SMPTE250M 디지털 영상세계를 보여준다. 그리고, 방송영상 망의 출력은 H.323 스테이션의 카메라 입력으로 보낸다. 음향도 유사한 형태로 라우팅 된다. H.323 시스템은 수동으로 관리될 수도 있고, PC Anywhere와 같은 제품을 이용하여 원격조정 될 수도 있다. 원격 조정의 경우, 원격조정 어플리케이션 자체는 H.323 단말 CPU상에 환영받지 못하는 부담이 있어 영상 성능을 받아들이지 못하는 범위까지 느리게 한다는 점을 인식해야 한다. 따라서, 원격조정은 H.323 호가 설정되기 전에 셋업 기능으로 만들어져야 하며, 호가 완료되었을 때는 다시 해제 기능이 있어야 한다. 그러나, 어떠한 공급업체도 이러한 종류의 게이트웨이를 구현할 계획은 아직 없다.

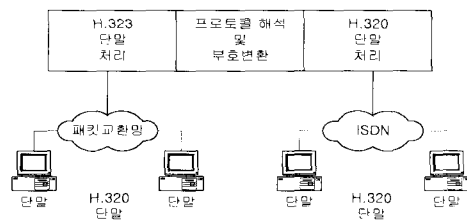


그림 3 H.323/H.320 게이트웨이

게이트웨이의 목적은 H.323 단말에서 비-H.323 단말까지의 특성을 반영하는 것으로, 아날로그와 디지털 음성 단말 링크, 비-H.323 단말 링크, 게이트키퍼 기능(의무 사항) 등이 주요 어플리케이션이다. 하나의 H.323 단말과 직접 통신할 수 있기 때문에, 다른 표준과의 접속이 필요 없다면 게이트웨이도 필요 없다.

단말은 H.245 및 Q.931 프로토콜을 이용하여 게이트웨이와 통신하는데, 게이트웨이를 통해 통신할 수 있는 실제 H.323 단말 수는 표준 회의 대상이 아니다. 또한, WAN 접속 수와 지원되는 독립적인 동시 회의 수는 제조업체가 결정할 일이다. 게이트웨이 기술을 H.323 규격에 접목시켜, ITU는 H.323을 전세계적인 영상 회의의 표준으로 자리잡게 만들었다.

다. 게이트키퍼(Gaetkeeper)

게이트키퍼는 기업 데이터망의 통합성을 유지하는데 필요한 주소제어, 접근제어, 대역폭제어, 영역관리, 호 제어 시그널링, 호 권한부여, 대역폭 관리, 호 관리, PBX 기능들을 수행한다.

다중점 회의를 지원하기 위해서 게이트키퍼는 점-대-점 회의에 참석한 두 터미널들로부터 H.245 제어 채널을 수신할 수 있다. 회의가 다중점 회의로 변경될 때, 게이트키퍼는 H.245 제어 채널을 MC로 양도할 수 있다. 게이트키퍼는 H.245 시그널링을 처리할 필요는 없고, 터미널간 또는 터미널과 MC간에 전달만 해주면 된다.

게이트웨이를 포함하는 LAN은 주소 변경(E.164 - 전송 주소)을 위해서 게이트키퍼도 포함해야 한다. 게이트키퍼를 포함하는 H.323 실체들은 같은 LAN상에 게이트키퍼를 포함하는 다수의 실체들이 있을 때 그들이 같은 지역(Zone)으로 구성될 수 있도록 하기 위해서 내부 게이트키퍼를 비활성화 시킬 수 있는 방법을 가지고 있어야 한다.

라. MCU(Multipoint Control Unit)

MCU는 3자 또는 그 이상의 단말간 회의를 지원한다. H.323 내에서 MCU는 MC(Multipoint Controller)와 MP(Multipoint Processor)로 구성되는데, MP는 없을 수도 있다. MC는 음향 및 영상처리를 위한 일반적인 용량을 결정하기 위해 모든 단말간 H.245 교섭을 조정하며, 또한, 음향 및 영상 스트림이 멀티캐스팅 될지를 결정하여 회의 자원을 제어한다. MC는 어떠한 매체 스트림도 직접 제어하지는 않으며, 음향, 영상 및 데이터 비트의 믹싱, 스위칭은 MP가 담당한다.

MCU는 회의의 통제센터와 같은 역할을 담당하여, 음성 활동이나 수동조정을 근거로 영상 소스를 스위칭 한다. 또한, MCU는 두 대의 상이한 시스템들이 함께 회의할 수 있게 해준다. 만일 사이트 A가 T1 직통접속을 가지고 있고, 사이트 B가 ISDN 성능을 가지고 있다고

하면, 그것들은 함께 직접 회의를 할 수 없다. MCU는 H.320 영상회의 표준 내에서 이 두 종류의 접속 사이에 브릿지 역할을 담당한다. T1 사이트는 직접 브릿지에 접속하고, ISDN 사이트는 브릿지로 다이얼할 것이다.

(2) 멀티포인트 회의

멀티포인트 회의는 중앙 집중식, 분산식, 혼합식 세가지 형태가 있다. 첫 번째는 중앙 집중식 멀티포인트 회의는 MCU를 사용하여 미디어, 데이터 스트림을 분배한다. 각 터미널은 MCU로 미디어 스트림을 전송하고, 선택되거나 혼합되어진 미디어 스트림을 선택하여 터미널로 되돌려 보낸다. 이것은 전통적인 H.320 MCU 능력과 유사하다. 두 번째 분산식 멀티포인트 회의는 각 터미널들이 멀티캐스트 또는 유사한 매커니즘을 사용하여 회의중인 모든 터미널로 미디어 스트림을 분배한다. 중앙의 MCU는 제거한다. 세 번째 혼합식 회의는 중앙 집중식과 분산식 회의의 두 개의 요소를 혼합한 것이다.

이들 회의 형태들을 지원하는 MCU는 MP와 MC로 구분된다. MP는 오디오 혼합, 비디오 혼합 또는 비디오 전환과 같은 미디어 처리를 수행한다. MC는 일반 통신 형태 그리고 미디어 채널 설정과 같은 회의 제어를 제공한다.

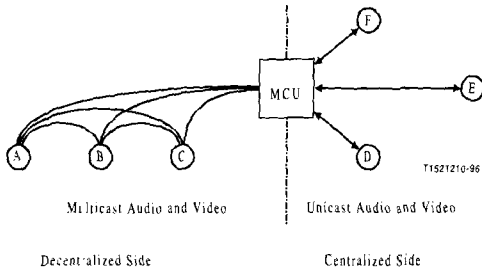


그림 4 혼합형 회의

(3) 호 모델 및 호 설정

H.323의 호 신호에 대한 두 개의 모델은 직접 호 신호방식과 gatekeeper를 경유한 호 신호방식이 있다. 이 때 gatekeeper는 종단(endpoint)들 사이의 모든 호 신호를 중계한다. H.323 내의 호 설정은 몇 개의 단계를 가진다. 첫째, gatekeeper를 사용하여 실행한다면, 종단의 호출은 gatekeeper로부터 호 허락을 요구할 때 H.225.0 ARQ(그림 5의 메시지 1)를 이용해야 한다. 만약 gatekeeper가 호를 허락한다면 gatekeeper는 ACF(2)로 응답한다; 달리, 요구를 거부할 때는 ARJ(2)를 사용한다. 만약 gatekeeper가 없거나, 또는 허락이 수신된 후, 호출하는 종단은 Setup(3) 메시지를 호출되는

종단으로 보낸다. 호출된 종단은 Call Proceeding(4) 메시지에 의해서 Setup(3) 메시지를 받았음을 알린다. 만약 호출된 종단이 들어온 호를 수락한다면, 위에서 설명한 ARQ(5)/ACF(6) 절차를 이용하여 gatekeeper로부터 우선 허락을 얻어야 한다. 허락을 수신한 후, 호출된 종단은 호출한 종단으로 Alerting(7)을 보내고, 이 과정은 들어온 호를 알리는 것이다. 만약 응답한다면, 호출된 종단은 Connect(8) 메시지를 호출한 종단으로 보낸다. 종단들이 H.245 제어 채널이 설정된 다음은 용량 교환이 시작되고, 오디오, 비디오 및 또는 데이터에 대한 미디어 스트림 채널이 열린다.

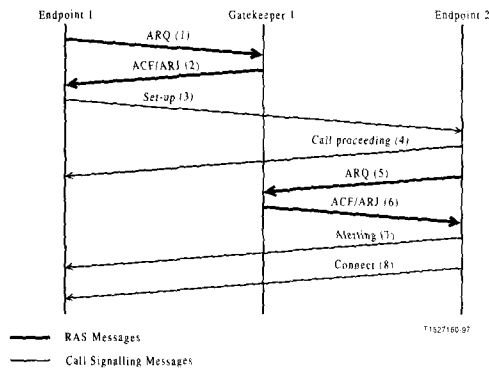


그림 5 호 설정 과정

위의 예는 직접 호 모델로 두 개의 종단은 같은 gatekeeper에 등록된 경우이다. 유사한 과정으로 gatekeeper 경유 모델과 멀티플 gatekeeper는 보다 복잡한 경우에 사용하곤 한다. 호 신호 과정은 H.225.0에 의해 맞추어진 것과 같은 Q.931이 기본이다. Q.931은 또한 ISDN 네트워크상의 H.320 터미널에 사용된다.

(4) 서비스 품질(QoS)

H.323은 LAN상에서 영상 전화 연결에 대한 QoS의 보장에 대한 어떤 매커니즘도 제공하지 않는다. 그러나, QoS 평가와 제어를 위한 다소의 도구를 제공한다. H.323의 제어와 데이터 채널은 TCP와 같은 신뢰성 있는 전송 서비스를 이용한다. 패킷들의 신뢰성 있는 전송의 제공은 수신을 승인하는 것과 패킷 손실시 재전송을 이용한다. 또한 오디오와 비디오를 위한 실시간 채널은 UDP와 같은 최고의 효과적 인 전송 채널을 이용하는데, 패킷의 신뢰성 있는 전송을 보장하지 않고 받은 것에 대한 승인과 재전송을 수행하지 않는다. 지터, 패킷 손실 그리고 네트워크 혼잡은 신뢰성이 없는 채널에서

불리한 영향으로 작용하여 QoS를 감소시킨다. H.323은 이것의 영향을 최소화하는 도구를 다음과 같이 제공한다.

H.245와 RTCP는 오디오와 비디오 패킷의 재전송을 요구하는 선택 매커니즘을 제공한다. 만약 지터 버퍼의 크기가 충분하고 전송시간이 짧으면, 이것은 손실 패킷을 재전송 할 수 있는 시간이 주어진다. 추가적으로 영상에 대한 복호기의 에러 은닉 기술은 정보의 손실을 효과적으로 최소화 할 수 있다.

Gatekeeper는 LAN상에서 H.323 호 위치의 수의 관리와 각 호의 대역폭 사용에 대해 네트워크 관리자의 설계자에 대한 매커니즘을 제공한다. 이것은 호에 대한 대역폭의 유효성을 보장하지 않고, LAN 트래픽상의 H.323 호의 충돌을 제어한다.

RTCP 송신자와 수신자보고는 미디어 스트림에 대한 송신자와 수신자 사이의 QoS 정보의 전송을 제공한다. 송신자보고는 송신자에서 수신자로의 정보를 제공한다. 이것은 오디오와 비디오 스트림 동기, 예상 데이터 비율, 예상 패킷 비율 그리고 송신자의 시간 간격에 대한 정보를 포함한다. 수신자보고는 수신자로부터 송신자로의 피드백을 제공한다. 이것은 손실 패킷 부분, 손실 패킷 누적, 수신된 마지막 순서 번호 그리고 상호도착 지터를 포함한다. 이 정보를 이용하여 네트워크 혼잡을 검출하고 미디어 스트림 데이터 비율 감소와 같은 상황 교정을 할 수 있다.

라우터(인터넷을 포함)를 포함한 네트워크는 예약 프로토콜을 지원한다. H.323 터미널은 이 예약 프로토콜을 우선하거나 호 설정 과정 동안에 사용할 수 있다. 네트워크 자원들은 라우팅 경로를 따라서 송신자로부터 수신자에게로 예약할 것이다. IETF RSVP(Resource Reservation Protocol)와 같은 예약 프로토콜은 라우터가 자원을 충분히 이용할 수 있는 QoS 요구를 지원하고, 사용자는 관리상의 허가가 예약을 만들기 위한 결정을 할 것이다.

III. 영상 압축 H.263

ITU-T에서 Video Coding for Low Bit Rate Communications라는 명칭하에 H.263 표준[3]을 제정했다. H.263은 원래 64Kbits/s이하의 저 전송률 영상통신을 위한 영상 부호화 표준으로 제시되었다. 그러나 현재는 저 전송률뿐만 아니라 폭 넓은 전송률에도 적용함으로써 경우에 따라서는 H.261[4]과 MPEG 부호화 방

식과도 경쟁할 수 있는 기법으로 인정되고 있다.

H.263의 부호화 기법은 H.261 기법을 기반으로 하지만 H.261의 비트열 구조의 일부가 선택사항이 되었고 약간의 추가적인 연산에 의해 저 전송률에서 보다 좋은 화질과 향상된 에러 복원을 제공할 수 있다. 따라서 ITU-T H.324, H.320, H.310과 같은 표준 비디오 단말 등에 적용되고 있다

H.263은 비디오 신호를 배경화면은 그대로 두고, 움직이는 부분만 압축하여 전송하므로 효율이 높고 그만큼 뛰어난 화질을 보상할 뿐만 아니라 화상 전송의 지연을 최소화하여 자연스러운 화질 보상을 할 수 있습니다.

H.263의 코딩 알고리즘은 H.261과 비슷하나, 성능 및 에러 복원을 향상시키기 위한 개선을 위해 Halfpel 벡터 연산 등의 기법이 도입되었고, 비제한적인 움직임 벡터(unrestricted motion vector), 조건부 산술부호화(syntax-based arithmetic coding), 고급예측(advanced prediction), 그리고 순방향과 양방향 예측을 사용하는PB-프레임(predictive and bi-directional predictive frame)과 같은 4개의 선택적 표준기법들을 사용함으로써 보다 향상된 압축 성능이 제공되었다. 이러한 추가 선택 사양을 통해 동일한 양자화 값에서 H.261보다 비트율은 절반 이하로, SNR은 2~3dB 정도 향상시킬 수 있다.

H.263은 화면의 크기에 대해 미리 정해놓고 있는데, 원래 출발은 CIF에서 시작했지만, 실제로 사용되는 크기는 1/4크기의 QCIF이다. H.261에서는 QCIF가 필수로 되어 있고, CIF가 선택 사양으로 되어 있었는데, H.263에서는 SQCIF와 QCIF가 필수로 되어있고, CIF,4CIF,16CIF는 선택사양으로 되어있다.

H.263은 H.261과는 달리 반화소 정밀도의 움직임 추정을 사용한다. 참고로 H.261은 한화소 정밀도의 움직임 추정을 사용한다. H.263이 반화소 정밀도의 움직임 추정을 사용하기 때문에 거의 두 배의 압축 효율을 얻을 수 있다고 한다.

H.263은 H.261 또는 MPEG처럼 움직임 보상을 이용한 예측 부호화와 DCT 변환에 의하여 시공간적으로 높은 상관성을 갖는 영상신호 성분들간의 상관성을 제거한 후, 양자화에 의하여 이 신호 성분들을 몇 개의 대표 값으로 표현함으로써 정보압축을 수행한다. 그리고 가변 길이 부호화(VLC: Variable Length

Coding)를 사용하여 통계적 상관성을 제거함으로써 보다 높은 압축 효율을 얻는다. 그림 6은 H.263 비디오 부호기의 구성을 보여준다. H.263 부호화는 H.261의 경우와 마찬가지로 입력 영상을 화면, GOB(Group Of Block), MB(Macro Block), 블록과 같은 계층적인 구조로 분할한 후 처리한다.

부호화기에 입력되는 각 영상은 K*16 라인으로 구성되는 GOB들로 분리된다. 여기서 K는 입력 영상 규격에 관계된 변수로 sub-QCIF, QCIF 그리고 CIF에 대해서는 K=1, 4CIF에 대해서는 K=2, 16CIF에 대해서는 K=4이다. 따라서 한 영상 당 GOB의 수는 입력 영상 규격이 sub-QCIF이면 6개, QCIF이면 9개, CIF와 4CIF 그리고 16CIF이면 18개이다. 각 GOB는 16*16인 휘도성분과 8*8인 색차성분들(CB, CR)로 구성되는 MB들로 구성되며, 각 MB들은 다시 4개의 휘도성분 블록과 2개의 색차성분 블록들로 구성된다.

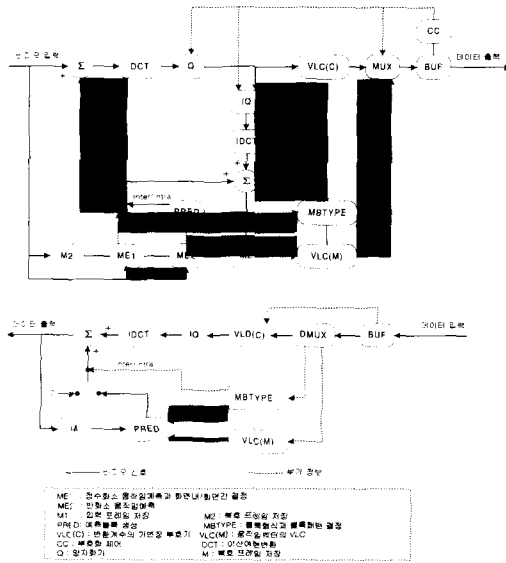


그림 6 H.263 부호기(위)와 복호기(아래)의 블록도

IV. 음성부호화 표준 G.711, G.722, G.723

음성부호화의 표준 G.711[5]은 64Kbps에서 약 3KHz까지의 전화급 오디오 품질을 제공하기 위하여 PCM 오디오 엔코딩과 미국, 유럽에서 주로 이용하는 μ -law 또는 A-law 방식을 사용합니다. G.722[6]는 G.711과 흡사하지만

64Kbps 대역폭에서 7KHz의 스테레오 오디오 품질을 지원합니다.

G.723[7]은 매우 낮은 비트율로 음성 또는 멀티미디어 서비스의 다른 오디오 신호 요소를 압축한다. 이 코더는 5.3Kbps 와 6.3Kbps의 두 비트율을 지원합니다. 높은 비트율은 더 좋은 질을 갖게된다. 낮은 비트율은 좋은 질을 제공하고 시스템 설계자들에게 추가적인 유연성을 제공한다. 이 코더는 복합의 제한된 양을 사용하고 있는 상기의 비트율들에 고품질로 음성을 표현하기 위해 최적화 되었다. 이것은 선형의 분석 대 통합 코딩을 사용하는 프레임에서 음성 또는 다른 오디오 신호들을 부호화한다. 높은 비트율 코더를 위한 자극 신호는 Multipulse Maximum Likelihood Quantization(MP-MLQ)이고 낮은 비트율 코더를 위한 자극 신호는 Algebraic Code Excited Linear Prediction(ACELP)이다. 프레임 크기는 30ms 이고 7.5msec의 추가적인 예비가 있다. 따라서 37.5msec의 총 알고리즘 지연을 초래한다. 이 코더에서 모든 추가적인 지연은 멀티플렉싱 프로토콜의 버퍼링 지연과 통신 선로에서 전송 지연과 실행의 처리 지연을 의미한다.

이 코더는 제한된 유연성을 사용하는 위의 비트율에서 높은 질을 갖는 목소리를 표현하는데 최대로 활용되어진다. 음악 또는 다른 오디오 신호들은 목소리만큼 충실히 표현되지는 않지만 이 코더를 사용하면 압축을 하고 압축을 풀 수 있다.

가. G.723의 인코딩

이 코더는 아날로그 입력의 전화 대역폭 필터링(ITU Rec.G.712)의 첫 번째 수행에 의해 얻어지는 디지털 신호와 동작되도록 설계되어진다. 그때 8000Hz로 샘플링되고 인코더에 입력을 위한 16bit 선형으로 PCM으로 변환된다. 디코더의 출력은 비슷한 방법으로 아날로그로 변환될 수 있다. 64kbit/s PCM 데이터를 위한 ITU G.711에 의해 서술되는 것과 같은 입/출력 특성은 디코딩 전에 16 bit PCM에서 적당한 형태로 또는 인코딩 전에 16bit 선형 PCM으로 변환되어질 수 있다. 이 권고안에서 인코더에서 디코더로 비트 스트림은 정의되어진다.

이 코더는 선형의 분석 대 종합 코딩의 원리를 기본으로 하고, 인지적으로 가중된 에러 신호를 최소화한다. 인코더는 각 240 샘플의 블록들(프레임들)로 동작한다. 그것은 8kHz 샘플링율에서 30msec와 같다. 각 블록은 DC 요소를 제거하기 위한 첫 번째로 높게 통과 필터되고 그때 60 샘플들로된 네 개의 서브 프레임들로 나뉜다. 모든 서브프레임에서 10번째 주

문 선행 예측 코더(LPC)필터는 처리되지 않은 입력 신호를 사용하여 계산되어진다. 마지막 서브 프레임을 위한 LPC 필터는 Predictive Split Vector Quantizer(PSVQ)를 사용하여 양자화되어진다. 비양자화된 LPC 계수들은 짧은 기간 지각있는 가중 필터를 구성하는데 사용되고 완전한 프레임을 필터하는데 인지적으로 가중된 음성신호를 얻는데 사용된다.

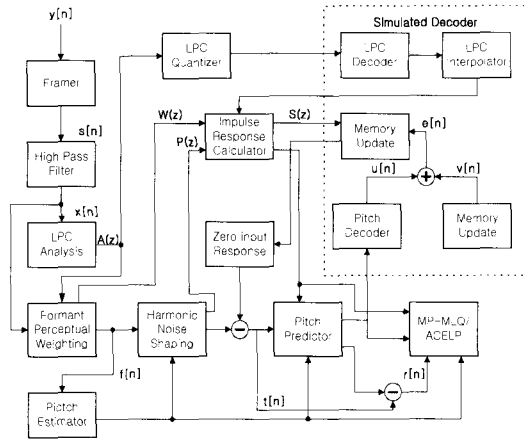


그림 7 G.723 부호화기

모든 두 개의 서브 프레임(120 프레임들)에서 개방 루프 pitch 주기(L_{OL})은 가중된 음성 신호를 사용하여 계산된다. 이 pitch 평가는 120 샘플들의 블록에서 평가된다. pitch 주기는 18에서 142 샘플들의 범위에서 연구되어진다.

이 점에서 음성은 기본 서브 프레임 당 60 샘플들에서 처리된다. 평가된 pitch 주기를 미리 계산함으로써, 조화된 잡음 모양을 이루는 필터가 구성된다. LPC 종합 필터, 지각있는 가중 필터 그리고 조화된 잡음 모양을 이루는 필터의 결합은 임펄스 반응을 만드는데 사용된다. 임펄스 반응은 더 많은 계산을 위해 그때 사용된다.

pitch 주기 평가를 사용함으로써 L_{OL} 그리고 임펄스 반응, 폐쇄 루프 pitch 예측치가 계산되어진다. 15번째 명령 pitch 예측치는 사용되어진다. pitch 주기는 개방 루프 pitch 평가 주위의 미분 값으로써 계산되어진다. pitch 예측치는 초기의 목표 벡터로부터 그때 공제된다. pitch 주기와 미분 값은 디코더에 전해진다.

마지막으로 자극의 비주기 요소는 근접해진다. 높은 비트율에서 multi-pulse maximum likelihood quantization(MP-MLQ) 자극은 사용되어지고 저 비트율에서 algebraic-code-excita-

tion(ACELP)이 사용된다.

나. G.723의 디코딩

압축된 음성신호가 프레임 단위로 구성되어 있으므로 복호화기에서의 처리도 역시 프레임 단위로 이루어진다. 그림 8에 전체 복호화기가 나타나 있다.

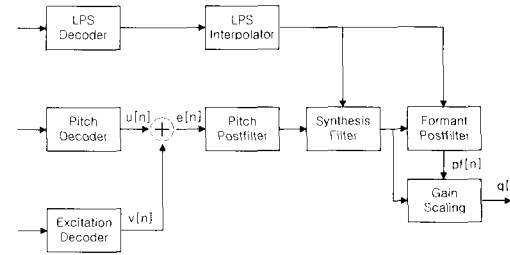


그림 8 G.723 복호화기

부호화기에서 부호화된 각 파라미터는 복호화기에서 사용하기 위해서는 다시 복호되어야 한다. 복호해야 할 파라미터로는 LSP, 피치주기, 여기신호가 있으며 이것들은 그림 8에서와 같이 복호화기를 거쳐서 합성음성을 만드는데 사용된다.

복호화기에서는 우선 양자화된 LPC 계수를 복호하여 LPC 합성 필터를 구성한다. 그리고 매 부프레임 마다 적응 코드북 여기신호와 고정 코드북 여기신호가 복호되어 합성 필터에 입력된다. 적응 후 필터는 포먼트와 forward-backward 피치 후 필터로 구성된다. 그리고 이득의 크기 조절은 포먼트 후 필터의 입력 에너지 레벨을 유지하도록 이루어진다. 여기서 피치 후 필터는 합성된 신호의 음질을 개선하기 위해 사용된다. 피치 후 필터에서의 신호처리는 매 부프레임 단위로 행해지며, 신호의 질을 개선하기 위한 방법으로는 피치주기의 배수에서의 SNR을 높이는 방법을 사용한다. 그리고 10차의 LPC 합성 필터는 앞에서 피치 후 필터링된 신호를 입력으로 하여 합성된 음성신호를 얻기 위하여 사용된다.

위에서 언급된 음성 부호화기는 프레임 손상에 강하게 설계되어 있다. 즉 잃어버린 프레임에 대하여 오류를 감추기 위한 처리방법이 복호화기에 포함되어 있다. 그러나 이러한 방법은 현재의 프레임이 손상된 것을 인식할 수 있을 때만 동작하게 되어 있으므로, 연속오류에 대한 대책은 되지만 불규칙한 비트 오류에 대한 오류 정정 방법은 아니다. 만일 잃어버린 프레임이 발생하면 복호화기는 원래의 복호처리 과정에서 오류 숨김 모드로 전환된다. 그리

고 LSP 계수나 여기신호에 관계없이 프레임 인터플레이션 과정이 수행된다.

나머지 인터플레이션은 잃어버린 프레임 이전의 프레임에 따른 두 가지의 다른 방법으로 진행된다. 우선 프레임을 voiced/unvoiced로 구분한다. 위에서 프레임을 두 가지로 구분하는 데는 최대 상호상관관계 함수가 사용된다. 즉 프레임의 마지막 120 샘플을 가지고 상호상관관계를 계산한다. 이때 최대 상관관계 값을 갖는 계수를 인터플레이션 할 후보 계수로 정한다. 그리고 가장 좋은 벡터의 예측이득을 조사하여 0.58dB 이상이면 voiced로, 그렇지 않으면 unvoiced로 구분한다.

만일 현재 프레임이 잃어버린 프레임으로 인식되고, 이전 프레임이 voiced로 구분되었다면 현재 프레임의 여기신호는 균일 불규칙 숫자 생성기를 이용하여 만들어진다. 그리고 현재 프레임이 voiced고 구분되었다면, 주기적인 여기신호를 이용하게 된다. 잃어버린 프레임이 다음에 두 개 더 계속되면, 새로이 생성된 벡터는 2.5dB 더 감쇄시킨다. 그리고 3개의 인터플레이션 프레임 후에는 출력이 완전히 묵음화된다.

V. 구현 및 실험

1. 영상 채팅을 위한 세션관리

H.323의 호제어를 바탕으로 본 시스템에서는 서버 데몬 프로그램에서 이를 처리하도록 구현하였다. 기본 메인 서버는 대화실 세션만 만드는 역할을 하며, 다음과 같은 사항들을 웹을 통해 사용자로부터 입력받는다.

(1) 기본 세션 서버 데몬

- 대화실 제목, 대화실 제한 인원, 대화실 종류를 사용자를 통해 입력받음
- 대화실 관리 DB에 개설된 대화실 정보 저장
- 대화실 전용 관리 서버 프로그램을 생성

대화실 전용 관리 서버 프로그램은 개설된 대화실을 실시간으로 관리하고, 새로운 참여자를 허가하며 이를 다른 참여자들에게 캐스팅해주는 역할을 한다. 그리고 기타 다음과 같은 기능을 수행한다.

(2) 세션 관리 서버 데몬

- 새로운 참여자 허가

- 허가된 참여자 대화실 DB에 첨가 후에 갱신
- 허가된 참여자 다른 참여자들에게 정보 전송
- 제한된 인원을 초과하지 못하도록 세션을 관리
- 참여자들 사이의 발언권 제어
- 참여자 퇴장 시에 제어신호를 다른 참여자들에게 전송
- 참여자들에게 일정간격으로 확인신호를 전송하여 연결을 유지

이러한 서버 프로그램을 바탕으로 웹을 통해 개설된 대화실 정보를 사용자들에게 보여주는 모듈을 동시에 제공한다. 클라이언트 프로그램을 웹 브라우저 플러그인으로 제공하지 않고 외부 프로그램을 구동시키는 플러그인을 제공하여 사용자의 편의를 고려하였다.

클라이언트 프로그램을 구동시키는 플러그인은 사용자가 선택한 대화실의 정보와 자신의 가입된 정보를 가지고 구동시킨다.

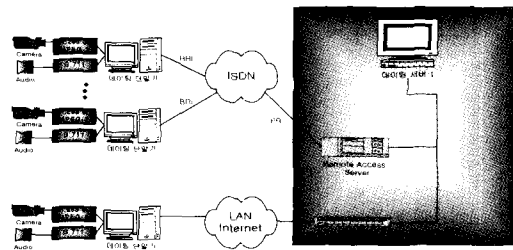


그림 9 멀티미디어 데이팅 시스템 구성도

(3) 클라이언트 프로그램의 기능

- 구동 시에 입력되는 정보를 가지고 세션 관리 서버에 참여요구
- 세션 관리 서버에서 허가신호를 받음
- 세션 관리 서버를 통해 기존의 참여자 정보를 수신
- 수신된 참여자 정보를 기반으로 클라이언트들 사이에 호 설정
- 클라이언트들 사이에 비디오/오디오/텍스트 호가 설정되면 클라이언트 프로그램에서 참여자 정보를 기반으로 RTP를 이용하여 다른 참여자들에게 자신의 정보를 캐스팅
- 일정간격으로 세션 관리 서버에 연결유지 신호를 전송하여 연결을 유지
- 새로운 참여자나 퇴장한 참여자를 실시간적으로 세션 관리 서버를 통해 수신하여 캐스팅 정보를 갱신

2. RTP/RTCP에서의 오디오/비디오

패킷 전송

(1) RTP/RTCP의 기본개념

이전의 RTP 라이브러리 설계와는 달리 본 개발에서는 트랜스포트 제어 메커니즘을 통합한다. 즉, 전송률 제어와 오류 제어를 RTP/RTCP 모듈에 통합한다. RTP는 1996년 IESG에 의해 RFC 1889로 채택되었고, 그 후 ITU-T 권고안 H.225.0의 일부가 되었다.

본 논문에서 구현한 RTP/RTCP 모듈에서 제공되는 기본 기능은 RTP 패킷 헤더 연산, RTP 페이로드 헤더 연산, RTCP 패킷 헤더 연산, 오류 제어와 RTP 계층 패킷 감시를 포함한다.

RTP/RTCP 모듈에는 RTP 흐름과 RTCP 패킷을 처리하고 관리하기 위한 요소들뿐만 아니라 다수의 부기(book-keeping) 요소들도 제공된다. 2대의 영상회의 단말이 사용할 미디어 유형에 동의하고 나면 그에 대응하는 RTP/RTCP 세션이 적당한 설정 정보(configuration information), 특히, 미디어의 페이로드 유형과 대역폭 요구를 가지고 개방된다. 이 설정 정보는 Bandwidth Manager와 Session Configuration Manager에 의해 계속 유지된다. 그리고 RTP/RTCP 모듈은 또한 대역폭이나 아니면 다른 시스템 자원이 불충분한 경우에 수락 제어(admission control)를 행한다 [8].

일단 RTP/RTCP 연결이 셋업되면 하층(lower layer) 네트워크 프로토콜 상에서의 정보는 Socket Agent에 의해 유지된다. 하나의 미디어 프레임이 RTP/RTCP 모듈에서 네트워크층으로 전송되면 Payload Flow Manager는 상층(upper layer)이 사전 협상된 트래픽 요구를 준수하도록 데이터 흐름을 감시하게 된다. 그런 다음 등록된 미디어 양식에 따라 Payload Packer는 미디어 데이터를 적당한 RTP 페이로드 양식으로 캡슐화한다. 그 다음에는 RTP Packer가 RTP 헤더를 준비된 페이로드에 첨부한다. 다음에 RTP 패킷은 Schedule Agent로 보내지고 Schedule Agent는 스케줄을 짜고 모든 활성화 RTP 세션을 전송한다.

한편, 네트워크에서 수신된 RTP 패킷은 먼저 RTP로 역다중화되며, RTCP는 각각 큐들을 수신한다. RTP 패킷은 캡슐화된 미디어 데이터를 얻기 위해 RTP와 RTP 페이로드 헤더들을 분리하도록 RTP Analyzer와 Payload Assembler로 전송된다. RTCP 패킷들은 일단 수신된 뒤 다음 분석을 위하여 RTP Agent로

보내진다. RTP Agent는 RTCP Analyzer의 로그 정보와 프로파일 정보에 근거하여 RTCP 송신기와 수신기 레포트를 작성한다. 네트워크 트래픽 관련 정보는 RTP/RTCP 모듈이 Traffic Reporter와 상호작용하여 호출기(caller)에 의해 얻어진다.

Profile Register의 관리를 통하여 이용자 정보와 이용자 서술(user-specified) RTCP 메시지가 커스텀화된다. 끝으로 Session Agent는 세션 정보를 관리하고, RTP세션 정보를 다른 모듈 정보에 제공하기 위해 RTP/RTCP 모듈의 기타 요소들과 통합된다.

(2) RTP/RTCP에서의 비디오 패킷 전송

카메라에서 비디오를 입력받아 코덱을 통해 압축되는데, 압축하면서 일정한 크기가 되면 이를 패킷화하여 대화 참가자들에게 이 패킷을 캐스팅한다. 이에 비디오에서는 H.263이나 MPEG-2의 압축특성상 이전 프레임이 전송되지 않으면 다음 프레임을 복호화 할 수 없기 때문에 정확한 전송이 요구된다. 그래서 본 시스템에서는 비디오 패킷에 대해서는 시간내에 완전한 전송을 보장하도록 설계되었다. 일정 패킷을 캐스팅한 후에 확인신호를 일정시간동안 기다리게 된다. 대화참가자 확인신호가 모두 확인된 후에 다음 패킷을 전송하게 된다. 비디오 전송속도는 대화참가자 중에 가장 좋지 않은 환경으로 대화중인 사용자에게 맞추어진다.

수신측에서는, 패킷 재조합 과정이 필요하므로 일정 크기의 버퍼를 사용하도록 하였다. 각 RTP 패킷 헤더에 타임스탬프와 시퀀스번호를 이용하여 패킷을 재조합하게 되는데, 일정시간 내에 요구한 패킷이 수신되지 않으면 현재 수신된 패킷들을 가지고 디코딩을 수행하게 된다 [9].

(3) RTP/RTCP에서의 오디오 패킷 전송

비디오와는 달리 오디오 데이터는 압축 특성상 일정한 시간적 종속성이 존재하지 않으므로 비디오 패킷 전송과는 달리 일정 패킷을 캐스팅한 후에 확인신호를 기다리지 않고 바로 다음 패킷을 대역폭에 비례하여 연속적으로 전송하게 된다.

수신측에서는 비디오 수신과 마찬가지로 일정한 크기의 수신버퍼를 사용하여 수신되는 오디오 패킷을 타임스탬프와 시퀀스번호를 이용하여 패킷을 재조합하게 되며, 만약 수신되지 않은 패킷이 있으면 G.711에 거채된 방법으로 오디오 에러 은닉을 수행하여 오디오 패킷들을

처리되도록 설계하였다.

3. 구현 결과

이 시스템은 클라이언트간에 직접 연결되어 큰 대역폭을 차지하는 비디오와 음성데이터를 주고받고 서버에는 단순히 세션을 관리하는 제어신호만을 전송하므로 서버에는 부하가 적게 걸리는 특징이 있다.



그림 10 구현된 시스템의 주화면

또, 영상과 음성의 비트율을 고려해 볼 때, 비디오는 평균 122Kbps의 비트율과 오디오는 64Kbps의 비트율이므로 전체적인 비트율은 186Kbps 이다. 하지만 오디오의 경우 항상 말을 하고있는 상태가 아니므로 평균적인 비트율을 생각한다면 비디오와 오디오의 비트율을 154Kbps 이내로 줄일 수 있을 것이다.

그림 10은 구현된 시스템의 주 화면으로 최대 10명이 대화에 참여 할 수 있고, 왼쪽의 화면은 텍스트를 통한 대화가 나타나고, 가운데는 음성채팅 상태를 나타내고, 오른쪽 화면은 대화실 상황을 나타내는 구조이다.

VI. 결론

본 논문은 기존 문자기반 채팅 서비스와 차별화된 고품질(음성/비디오 사용, 초당 15 프레임 이상 채팅 참가자들의 비디오 동시 제공)의 멀티미디어 데이팅 시스템을 개발하기 위한 솔루션으로, 네트워크를 통해 5인 이상이 참여하는 데이팅에서 참가자가 실시간으로 비디오/오디오/텍스트 데이터들을 주고 받을 수 있도록 설계되었다. 기존의 방법과는 달리 본 시스템은 영상회의의 표준인 H.323에 근거하여 최대한 따르려고 노력하였으며, 실질적으로 가장 중요한 비디오/오디오 코덱을 권고안에 따라 사용하였다.

서버의 과부하 문제를 고려하여 중앙 집중

형의 서비스를 배제하고 제어신호 및 기타 세션관리 등은 서버에서 이루어지나, 비디오/오디오/텍스트 데이터는 참가자들이 서로간의 네트워크를 이용하여 주고받을 수 있도록 설계하였다.

본 시스템을 이용하여 다지점간의 영상회의나 원격 화상 교육 그리고 본 시스템의 목적인 다자간 멀티미디어 데이팅 서비스를 수행할 수 있다.

참고문헌

- [1] ITU-T Recommendation H.323, Visual telephone systems and equipment for local area networks which provide a non-guaranteed quality of service, Geneve, 1996.11
- [2] ITU-T Recommendation, T.120, Data protocols for multimedia conferencing.
- [3] ITU-T Recommendation, H.263 Video Coding for low bit rate communication, 1998.10.
- [4] ITU-T Recommendation H.261, Video Codec for Audio Services at $p \times 64$ kbit/s, 1993.
- [5] ITU-T Recommendation, G.711, Pulse Code Modulation(PCM) of Voice Frequencies. 1993.
- [6] ITU-T Recommendation, G.722 7kHz AUDIO WITHIN 64KBIT/S
- [7] ITU-T Recommendation G.723.1, Dual rate speech Coder for multimedia Communications Transmitting At 5.3 and 6.3kbit/s, 1996.03
- [8] Network Working Group, RFC 1890, RTP Profile for Audio and Video Conferences with Minimal Control, 1996
- [9] Network Working Group, RFC 2032, RTP Payload Format for H.261 Video Streams, 1996

저자소개

박 한 업



- 1997년 2월 군산대학교 정보통신공학과(공학사)
- 1999년 2월 군산대학교 대학원 정보통신공학과(공학석사)
- 2000년 3월 - 현재 군산대학교 대학원 정보통신공학과 박사과정
- 1998년 - 현재 (주) 파스칼 소프트웨어 개발팀장

• 관심분야 : 영상 통신, 비디오 코딩

나 인 호



- 1998년 2월 울산대학교 전자계산학과(공학사)
- 1991년 2월 중앙대학교 대학원 전자계산학과(공학석사)
- 1995년 8월 중앙대학교 대학원 전자계산학과(공학박사)
- 1995년 9월~현재 : 군산대학교 전자정보공학부 조교수

- 1997년 7월~현재 : 전주 첨단 영상산업 추진협의회 위원
- 1997년 7월~현재 : 한국 해양정보통신학회 편집위원
- 1997년 12월~1998년 3월 : 전자통신연구원 초빙연구원
- 1999년 10월~2000년 10월 : 전주국제컴퓨터게임축제 조직위원
- 관심분야 : 멀티미디어 통신시스템, 분산시스템, 병렬처리

최 연 성



- 1982년 2월 중앙대학교 전자공학과 졸업(공학사)
- 1984년 2월 중앙대학교 전자공학과 대학원 졸업(공학석사)
- 1990년 2월 중앙대학교 전자공학과 대학원 졸업(공학박사)

- 1988년 3월 - 1991년 2월 제주대학교 정보공학과 조교수
- 1991년 6월 - 현재 군산대학교 전자정보공학부 부교수
- 1995년 - 1996년 군산대학교 전자계산소장
- 2000년 9월 - 현재 통일 IT 포럼 창립 회원(전자신문)
- 관심분야 : 영상 처리, 멀티미디어 시스템, 신호 처리 시스템