

H.323을 이용한 다자간 영상 채팅 시스템 개발

나형용^{*} · 김상길^{**} · 최영식^{**} · 이상홍^{**} · 최연성^{*} · 박한엽^{*} · 신현숙^{*} · 김동호^{*} · 김동욱^{***}
^{*}군산대학교, ^{**}한국통신 멀티미디어연구소, ^{***}전주대학교

Developing of Multipoint Video Chatting System using H.323

Hyoung-yong Na^{*} · Sang-gil Kim^{**} · Young-sik Choi^{**} · Sang-hong Lee^{**} · Yeon-sung Choi^{*} ·

Han-yeop Park^{*} · Hyun-sook Shin^{*} · Dong-ho Kim^{*} · Dong-wook Kim^{***}

^{*}Kunsan National University, ^{**}Korea Telecom Multimedia Tech.Lab., ^{***}Jeonju University

E-mail : poleon90@kornet.net

요 약

영상 채팅은 기존의 문자 위주의 채팅에서 벗어나 멀티미디어 시대에 맞춘 영상과 음성을 지원할 뿐만 아니라, 파일을 전송할 수도 있는 멀티미디어 기반의 채팅 시스템이다. 특히, 이 멀티미디어 채팅 시스템을 통해 화상통화를 할 수 있고, 또한 국제 전화 및 시외전화를 사용할 수 있다. 본 논문에서는 기존 문자기반 채팅 서비스와 차별화 된 고품질의 멀티미디어 데이팅 시스템을 개발하여, 네트워크를 통해 5인 이상이 참여하는 데이팅에서 참가자의 음성과 문자는 실시간 전송된다. 초당 15 프레임 이상 비디오를 동시에 송수신 할 수 있다. 기존의 방법과는 달리 본 시스템은 ITU의 영상회의 표준인 H.323에 근거하여 수정하여 사용하였으며, 실질적으로 가장 중요한 비디오/오디오 코덱은 권고안에 충실히 따랐다.

1. 서 론

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 모뎀 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 포

맷 정의, 공유 정보 처리 기술 등이 포함되어 있으며, 이종의 영상 회의 시스템 간의 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.

II. 영상회의 또는 채팅시스템 개발 동향

1. 다중 다자간 영상채팅 현황

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 결합인 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 회의입장, 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.

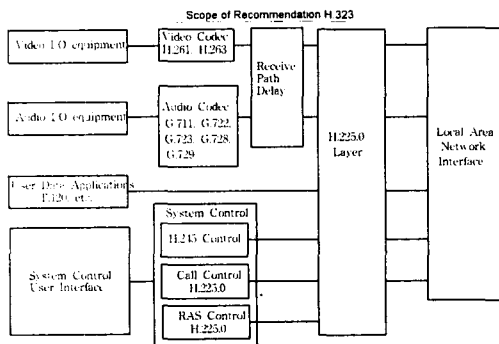


그림 1. H.323 시스템 구성도

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.

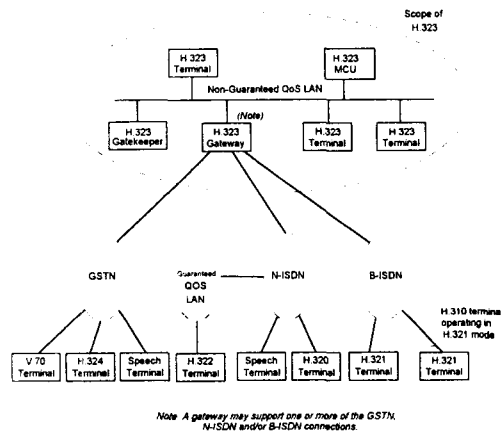


그림 2 H.323의 연결구성도

H.225.0은 호 신호, 회의 등록, 입장뿐만 아니

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

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 폭넓은 전송률에도 적용함으로써 경우에 따라서는 H.261[4]과 MPEG 부호화 방식과도 경쟁할 수 있는 기법으로 인정되고 있다.

H.263의 부호화 기법은 H.261 기법을 기반으로 하지만 H.261의 비트열 구조의 일부분이 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 보낼 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, 실제적으로 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 선택 사양으로 되어있다.

H.263은 H.261과는 달리 반화소 정밀도의 음

직접 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.

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 인코딩과 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 지원한다. 높은 비트율은 더 좋은 질을 갖게 된다. 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 30ms이고 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.

V. 구현 및 실험

제 1절 영상 채팅을 위한 세션관리

H.323의 호제어를 바탕으로 본 시스템에서는 서버 데몬 프로그램에서 이를 처리하도록 구현하였다. 기본 메인 서버는 대화실 세션을 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 캐스팅 해주는 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.

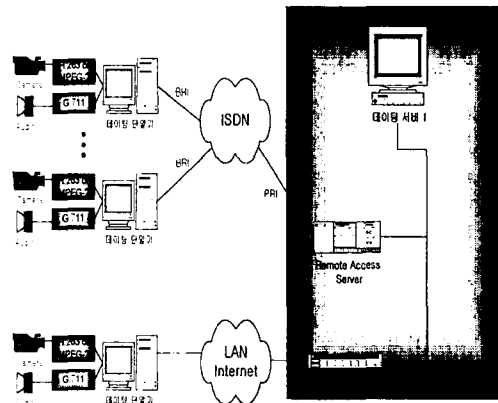


그림 3. 멀티미디어 데이팅 시스템 구성도

- (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

제 2 절 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].

(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 확인 신호를 일정시간동안 기다리게 된다. 대화참가자 확인신호가 모두 확인된 후에 다음 패킷을 전송하게 된다. 비디오 전송속도는 대화참가자 중에 가장 좋지 않은 환경으로 대화중인 사용자에게 맞추어 진다.

수신측에서는, 패킷 재조합 과정이 필요하므로 일정 크기의 버퍼를 사용하도록 하였다. 각 RTP 패킷 헤더에 타임스탬프와 시퀀스번호를 이용하여 패킷을 재조합하게 되는데, 일정시간내에 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 중속성이 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 거제된 방법으로 오디오 에러 은닉을 수행하여 오디오 패킷들을 처리되도록 설계하였다.

제 3 절 구현 결과

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

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 4는 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.



그림 4. 구현된 시스템의 주화면

VI. 결 론

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 영상회의 표준인 H.323에 근거하여 최대한 따르려고 노력하였으며, 실질적으로 가장 중요한 비디오/오디오 코덱을 권고안에 따라 사용하였다.

서버의 과부하 문제를 고려하여 중앙 집중형의 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 $p \times 64\text{ kbit/s}$, 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