

분산 구조를 위한 텔레프레즌스 클라이언트 구현 방법

황인기, 강현식, 김도영

한국전자통신연구원

ikhwang74@etri.re.kr, gadin@etri.re.kr, dyk@etri.re.kr

Implementation of telepresence client system for distributed network architecture

Inki Hwang, Hunsik Kang, Doyoung Kim

ETRI

요 약

텔레프레즌스 시스템은 기존의 영상회의 시스템과 비교하여, 라이프 사이즈의 출력, 고품질의 영상 및 음성 출력을 가능하게 함으로써 참석자로 하여금 실제 회의에 참석하고 있는 듯한 몰입도를 제공하여 원거리의 출장을 대체할 수 있는 시스템이다. 기존의 텔레프레즌스 시스템은 각 클라이언트의 미디어 정보를 중앙 서버에서 취합, 재구성하여 각각의 클라이언트로 전달하는 구조를 갖는다. 이러한 구조는 기능의 구현은 용이하나 모든 트래픽이 서버로 집중되어 서버가 설치되어 있는 통신 노드에 부하가 집중되는 단점을 갖는다. 이러한 문제점을 해결하기 위해 중앙 서버의 호처리 기능과 미디어 처리 기능을 통신망에 분산시키는 구조가 분산형 텔레프레즌스 구조이며, 분산 구조 하에서 각각의 클라이언트는 기존의 중앙 서버에서 처리되었던 미디어의 재구성 기능, 텔레프레즌스 서비스를 위한 고품질의 비디오 및 오디오 처리 기능, 협업을 위한 문서 공유 기능 등이 요구된다. 본 논문에서는 분산형 구조에서 요구사항을 만족하는 텔레프레즌스 클라이언트 시스템의 구현 방법에 대하여 기술한다.

1. 서론

과거 한 사무실에서 얼굴을 맞대며 업무를 처리하던 시대를 지나 현재는 세계 곳곳에 위치한 동료들과의 긴밀한 협조와 고객들과의 만남이 필요한 시대가 도래하였다. 전세계에 있는 동료, 고객들과의 만남을 위한 출장은 시간과 경제적인 손실을 야기시켰고, 이러한 문제를 해결하기 위해 제안된 방법이 화상회의 시스템이다.

화상회의 시스템의 요구 사항은 품질, 조작 편의성, 그리고 신뢰성이 있으며, 이 중 품질 요소는 통화자로 하여금 상대방과 대면하여 회의를 하고 있는 듯한 느낌을 줄 수 있도록 고품질의 화상/음성을 제공하고, 회의 참석자들의 협업을 위한 프레젠테이션 등의 서비스 지원이 있어야 한다. 몇몇의 회사들에서 텔레프레즌스라는 보다 고품질의 화상회의 시스템을 시장에 선보였다[1]. 하지만 이러한 시스템은 각 클라이언트의 미디어 정보를 중앙 서버에서 취합하여 처리하는 구조를 갖고 있기 때문에 모든 트래픽이 서버로 집중되는 문제점을 갖는다. 이러한 문제점을 해결하기 위해 제안된

방법이 분산형 텔레프레즌스 구조[2]이며, 본 논문은 이러한 분산형 텔레프레즌스 구조를 위한 클라이언트 구현 방법에 대하여 기술하고자 한다.

본 논문의 구성은 다음과 같다. 2 절에서는 분산형 텔레프레즌스 클라이언트 시스템의 구조에 대하여 기술한다. 3 절에서는 몰입도를 향상시키기 위한 기능들에 대하여 기술하며, 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 분산형 텔레프레즌스 클라이언트 구조

그림 1 은 분산형 텔레프레즌스 클라이언트 시스템의 구조를 보여준다. 클라이언트 시스템은 크게 미디어 엔진 블록, 네트워크 엔진 블록, 그리고 응용 엔진 블록으로 구성된다.

미디어 엔진 블록은 비디오 및 오디오 신호를 네트워크에 전달하기 위한 부호화 및 복호화 기능을 담당한다. 비디오 신호의 경우, 카메라 입력 신호를 부호 및 복호에 적합하도록

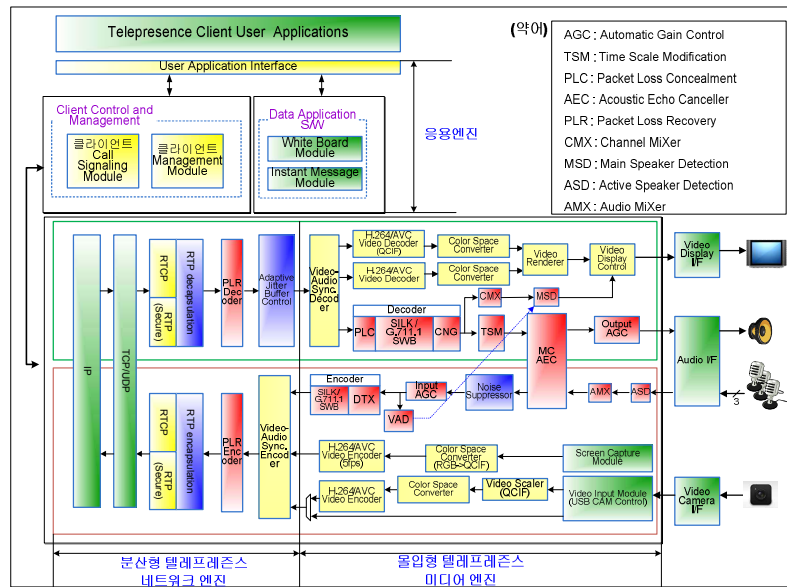


그림 1 분산형 텔레프레즌스 클라이언트 구성도

컬러공간 변환기(Color Space Converter)에서 형식을 변환한 후 부호 또는 복호 처리를 한다. 슈퍼 와이드 밴드급의 오디오 신호는 품질 향상을 위해 어커스틱 반향신호 제거, 자동이득 조절기 및 잡음 제거기 (Noise Suppressor)를 거친 후 부호화를 진행한다. 기존 영상회의의 시스템이 동질의 다자 화면을 제공하는 것과 달리, 구현된 미디어 엔진에서는 Full-HD 급 동영상과 동시에 작은 화면의 QCIF 급의 비디오를 전송하는 듀얼 비디오 스트림 전송방식을 사용한다. 듀얼 비디오 스트림을 사용하는 경우, 고화질의 비디오 영상을 요구 받지 않은 클라이언트는 고품질 영상을 네트워크 상으로 전달할 필요가 없어 네트워크의 부담을 줄일 수 있고, 영상을 재생하는 클라이언트에서는 고품질 영상의 많은 데이터 량을 처리할 필요가 없어 프로세서의 부담을 덜어주게 되는 장점을 갖는다. 분산 구조에서는 서버에서 각 클라이언트의 영상을 재구성하지 않기 때문에, 각 클라이언트는 회의에 참석 중인 다른 클라이언트로부터 전달된 영상 정보를 조합하는 기능을 수행해야 한다. 한 시점에서 하나의 고화질 영상과 복수개의 QCIF 영상을 복호하여 화면을 구성해야 하며, 사용자의 요구에 따라 위치를 변경하거나 고화질 영상의 대상을 바꾸는 기능을 제공한다.

듀얼 비디오 스트림을 위해 카메라로부터 입력되는 신호 중 QCIF 급 비디오 생성은 소프트웨어 코덱으로, 고화질의 HD 급 비디오 생성은 하드웨어 인코더를 이용하여 부호화한다. 영상 신호의 부호화 방식은 H.264 AVC 를 이용한다[3]. 이들 영상 신호는 움직임의 부드러움을 제공하기 위해 30 프레임 율 (frame rate)을 사용하며 오디오 신호는 20ms 주기의 음성 프레임을 0~14 KHz 의 대역폭까지 부호 및 복호 가능한 슈퍼 와이드 밴드급 음성코덱인 G.711.1 SWB[4] 또는 SILK[5]를 통해 상위계층으로 전달된다.

네트워크 엔진 블록은 트랜스포트 및 네트워크 품질 향상을 위한 종합적인 기능을 제공한다. 트랜스포트 엔진으로 전달되는 영상신호는 NAL (Network Abstraction Layer) 형식을 가지며, 네트워크 전송 프로토콜에 맞게 1,500 바이트 크기를 가지는 작은 조각(Fragment)으로 분할된다. 이렇게

분할된 조각들은 UDP (User Datagram Protocol) 패킷화 되고 RTP/RTCP 에 캡슐화되어 네트워크를 통해 전달된다. 네트워크 대역폭 상황이 달라지는 경우에는 네트워크 품질 성능지표(Network Quality Performance Metric)를 통해 성능단계별 임계치와 비교하여 전송대역폭을 조절한다. 전송대역폭의 조절은 H.264 AVC 코덱의 비트 율의 재조정, 전송 프레임화 비율의 조정 및 화면해상도 변경을 통해 가능하다. 오디오 신호의 경우 전송 대역폭이 크지 않기 때문에 적응대역폭 조절을 통하지 않고 수신신호의 패킷 손실 율을 구하여 그에 따른 복구 방식을 적용하는 지능형 패킷 손실 복구를 이용하여 네트워크 망의 시변성에 대응하여 끊기지 않는 실감 오디오 텔레프레즌스를 가능하게 한다.

응용 엔진 블록은 텔레프레즌스 사용자 인터페이스 및 협업 서비스 등의 응용서비스를 제공한다. 사용자의 편리성을 증진하기 위해 사용자의 디스플레이 장치 개수에 따른 다중 레이아웃을 제공하는 다중 스크린모드 레이아웃을 사용한다. 최대 3 개의 스크린을 제공하며 사용자의 의향에 따라 화상회의 시스템에서의 제공서비스 화면을 달리하여 사용자의 몰입도를 향상시키는 서비스를 제공한다. 텔레프레즌스 서비스의 협업 기능을 제공하기 위해 워드, 한글 문서, pdf 등 다양한 종류의 문서를 공유할 수 있는 기능, 프리젠테이션 기능, 원활한 화상회의의 진행을 위한 화이트보드 및 인스턴트 메시지 기능을 제공한다.

표 1 은 구현된 클라이언트 시스템의 사양이다.

표 1 클라이언트 사양

주요 기능 및 기술	주요 특징
듀얼 비디오 스트림 송수신	주화자(Main Speaker) : Full-HD / HD 급 영상 전송 회의참석자 : QCIF 급 영상전송
전송대역폭	3~5Mbps (Full-HD) 1.5~2Mbps (HD), ~200Kbps (QCIF)
미디어 코덱	비디오 : H.264 AVC

	오디오 : G.711.1 SWB, SILK (광대역, 슈퍼 와이드 밴드 코덱 선택 가능)
비디오 특징	다중 스크린 모드 지원 및 자동화면 배치 고해상도 지원 : 1920x1080p, 1280x720p
협업기능	문서공유, 화이트보드, 인스턴트 메신저
몰입도 향상 기능	어커스틱 에코제거 기능 주화자 및 능동화자 검출기능
미디어 품질제어	음성: PLR 영상: 채널상태 검출 제어

3. 몰입도 향상을 위한 기능

몰입도 향상을 위한 기능은 텔레프레즌스 참석자로 하여금 보다 회의에 몰입할 수 있도록 도와주는 기능을 의미하며, 구현된 텔레프레즌스 시스템에서는 몰입도 향상을 위해 다채널 어커스틱 에코 제거 기능과 주화자 검출 기능을 제공한다.

다채널 어커스틱 에코 제거 기능은 스피커 출력 신호가 다양한 경로를 통하여 세 개의 마이크로 다시 입력되어 발생하는 다채널 어커스틱 에코를 제거하기 위한 기능으로 PBFLMS (Partitioned Block Frequency-domain Least Mean Square) 방식의 적응형 필터를 이용하여 빠른 시간 내에 수렴이 가능하도록 구현하였으며, 2 EPM (Echo Path Model) 구조를 이용하여 더블 토크나 갑작스런 주변 환경 변화로 인해 발생할 수 있는 적응형 필터의 계수 값 발산을 막을 수 있도록 구현되었다. 어커스틱 에코의 후처리를 위해 감쇠된 에코의 에너지 레벨에 적응하는 노이즈 제거 장치를 두어 어커스틱 에코 제거 장치의 성능을 향상시킬 수 있도록 구현하였다.

주화자 검출 기능은 텔레프레즌스에 참석하고 있는 다수의 참석자 중에서 현 시점에 회의의 주도권을 쥐고 있는 참석자를 판단하여 다른 참석자들의 고화질 화면을 자동으로 해당 참석자로 변경시켜 주는 기능이다. 이는 텔레프레즌스 회의 중 발표자의 화면을 고화질로 보기 위해 사용자가 수동으로

화면을 조작하는 불편함을 해소하여 참석자로 하여금 보다 회의에 몰입할 수 있도록 도와준다.

4. 결론

본 논문에서는 분산 구조를 갖는 텔레프레즌스 시스템에서 클라이언트를 구현하기 위한 방법에 대하여 기술하였다. 구현된 텔레프레즌스 클라이언트는 Full-HD 급 고화질 영상과 QCIF 급의 참석자 영상을 듀얼로 제공하고, 슈퍼 와이드 밴드급의 오디오를 제공하며, 한 시점에서 하나의 고화질 영상과 다수의 참석자 영상을 동시에 보여준다. 협업 기능을 위하여 문서 공유, 화이트 보드, 인스턴트 메시지 기능을 제공하고, 사용자 편의에 따라 최대 3 개의 스크린을 자유로이 레이아웃하여 사용할 수 있게 구성되었다.

또한 텔레프레즌스의 몰입도를 향상시키기 위하여 다채널 어커스틱 에코 제거 기능과 주화자 검출 기능을 제공한다.

본 연구는 미래창조과학부가 지원한 2013 년 정보통신·방송 (ICT) 연구개발사업의 연구결과로 수행되었음

참조문헌

- [1] "Cisco Telepresence System", 2010, <http://www.cisco.com>
- [2] SH choi, "The design of participant-centric distributed telepresence architecture", ICEIC 2013
- [3] Jörn Ostermanns et al., " Video coding with H.264/AVC: Tools, Performance, and Complexity," IEEE Circuits and Systems Magazine, First Quarter 2004, pp. 7-28.
- [4] ITU-T G.711.1 Annex D, " Superwideband extension of G.711.1", 2008
- [5] <https://developer.skype.com/silk>

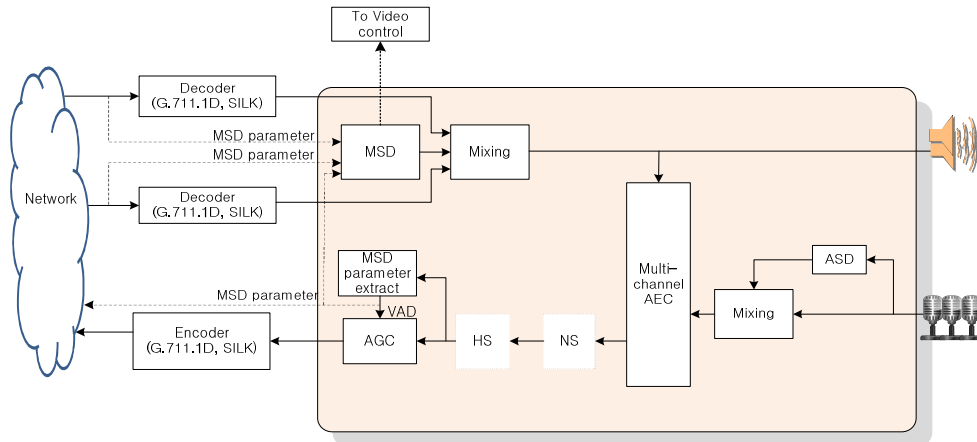


그림 2. 몰입도 향상을 위한 오디오 엔진 구조도