부가 영상 삽입기능을 갖는 화상회의 서버 구현

권성환(權聖煥), 성영락(成榮洛), 오하령(吳夏嶺)
국민대학교 전자공학과
전화 : (02) 910-4645 / 팩스 : (02) 910-7586

Implementation of a Video Conference Server with Additive Video-Stream

Seong-Hwan Kwon, Yeong-Rak Seong, Ha-Ryoung Oh
Kookmin University Electronic Engineering Department
E-mail : hwaney@zeus.kookmin.ac.kr, yeong@kmu.kookmin.ac.kr, hroh@kmu.kookmin.ac.kr

Abstract

In this paper, we implement a video conference system, which connects remote PCs. Each system receives video and audio data from devices attached to a PC and compresses them by using CODECs. It sends the compressed data to the opponent. Compared to the existing ones, the proposed system can combine two video streams into a new one. Thus, additional video data can be transferred to the opponent during a call.

1. 서론

최근 들어 인터넷과 고속통신망이 급속도로 발전하고, IMT-2000이 차세대를 이끌 통신망으로 각광 받고 있다. 이는 전체적인 네트워크 대역폭이 커지는 결과를 가져오며, 따라서 용량이 큰 멀티미디어 데이터의 실시간 송수신이 가능해 진다. 멀티미디어 데이터와 큰 네트워크 대역폭의 연계로 영상전화, 화상회의 및 영상을 이용한 다양한 부가 서비스들이 창출될 수 있다. 본 논문에서는, 멀티미디어 정보의 상호 운용성을 높이기 위한 시도의 일환으로, 네트워크로 연결된 두 대의 PC 양단간에 음성 데이터, 영상 데이터를 전송하는 화상통화 시스템을 구현한다. 기존의 화상 통화 시스템과는 달리 제안한 시스템에서는, 사용자의 영상에 각종 정보 및 광고 영상 등의 부가 영상을 결합하여 통신한다[1]. 본 논문에서는 이를 구현하기 위해 2장에서 시스템 구현에 참고한 표준안, 구현 환경 등에 대해 간략히 설명하고, 부가영상의 대치에 대해 살펴본다. 3장에서는 제안한 시스템을 구현하기 위한 설계안을 제시하고, 4장에서 이를 구현한다. 5장에서는 결론과 향후 연구방향에 대해 제시한다.

2. 관련 연구

2.1 표준안

화상전화에서 필요로 하는 멀티미디어 데이터는 주로, 음성·영상 데이터로 이루어지며, 이들을 단순 PCM(Pulse Code Modulation)으로 처리할 경우, 과다한 데이터 양으로 인해 저장이나 전송에 어려움이 생기게 된다. 따라서, 멀티미디어 시대와 함께 대두되는 화상통화의 핵심은 음성·영상 데이터의 효율적인 압축이라고 할 수 있다[2].

현재 사용되고 있는 화상전화의 표준안은, ISDN이나 고속 디지털 전용회선에서의 화상통화에 대한 ITU-T의 권고 내용인 H.320, 패킷 기반 네트워크(LAN)에서 실시간 멀티미디어 통신과 화상회의를 구현하기 위한 표준인 H.323, PSTN에 기반 한 H.324의 3가지 권고안으로 크게 나눌 수 있다. 각 표준안은 세부 하위 표준안을 포함하며, 세부 하위 표준안은 각각 영상 및 음성 데이터 전송, 이들간의 멀티플렉싱 등에 대한 표준 규격을 규정한다. 세부 하위 표준안은 음성 데이터, 영상 데이터에 관련된 표준 압축 방식이며, 제안한 시스템을 구현하기 위해 사용한 영상 압축 방식은 최근 화상통화에서 많이 사용되고 있는 영상 압축 표준은 H.263이다. H.263은 저속 통신망(아날로그 전화망, 무선망)을 이용한 오디오/비디오 서비스에서 동영상정보를 압축하기 위한 표준이며, 저속 통신망에서의 멀티미디어 서비스를 위한 제어 표준인 H.324 등에서 사용되고, 이에 대응하는 음성압축표준은 G.723.1이다[3].

2.2 윈도우 멀티미디어 시스템

본 논문에서 제안하는 시스템은 윈도우 멀티미디어 시스템과 밀접하게 연관되어 수행된다. 윈도우 멀티미디어 시스템은 시스템에 장착된 멀티미디어 장치들과

이들을 제어하는 드라이버와 함께 비디오 코덱, 오디오 코덱 등의 멀티미디어 구성 요소들을 제어, 관리한다. 아울러, 윈도우 멀티미디어 시스템은 기본적으로 지원되는 코덱 외에도 제 3자에 의해 개발되어 배포될 수 있는 표준 형식을 제안하고 지원하기 때문에, 표준 규격에 따르는 서비스가 윈도우 시스템에서 제공될 수 있는 기반을 마련해 준다. 본 논문에서 사용되는 윈도우 멀티미디어 시스템의 구성 요소로는 영상 입력, 음성 입출력, 영상 코덱, 음성 코덱 등이 있다.

2.3 부가영상대치

본 논문에서 제안된 부가영상대치는 H.261 영상 스트림을 이용해 선행되었다. 이에 대한 내용은 「H.261 비디오 스트림상의 부가영상 삽입 알고리즘의 개발」[1]에 근거한다.

기술한 논문에서는 H.261 영상 전체를 부호화하는 대신 실제 필요한 부분, 즉 부가영상이 대치되는 부분만 부호화하고 복호화하는 방법이 제안되었다. 제안한 방법에는 움직임 보상 정보가 영상에 아무 영향을 미치지 않을 경우에는 허프만 복/부호화만을 사용하여 부가영상을 단순 대치하고, 움직임 보상 정보가 부가영상과 겹치게 되어 부가 영상이 포함된 영상을 전달받는 측에서 문제가 될 경우만 미리 복호화 해 두었던 영상데이터를 인트라 블록으로 부호화하여 전달하는 과정이 포함되어 있다.

상기 논문에서 제안된 알고리즘은 H.261 영상 입력을 사용하는 기존 화상통화 단말기를 위해 제안되었다. 본 논문에서는 위 알고리즘이 일반 PC 상에서 적용될 수 있는 환경을 구현한다. 따라서, 영상 데이터의 경우 PC 환경을 기준으로 하는 화상통화에 적합한 H.263 규격을 사용한다. 아울러, 본 논문에서 구현하는 시스템에서는 원 영상의 부호화 이전 단계에서 부가영상을 대치하는 방법을 사용한다.

3. 시스템 설계

제안한 시스템의 설계는 영상 데이터 입출력, 음성 데이터 입출력, 부가영상대치, 마지막으로 입출력 데이터를 규정하는 순서로 진행하였다.

3.1 데이터 입출력

제안한 프로그램의 실행은 그림 1과 같이 영상 입력 장치의 유무를 확인하는 과정을 수행함으로써 시작된다. 영상 입력장치가 장착되어 있을 경우, 입력장치에 대한 제어권을 획득하고, 사용자의 영상을 입력 받는다. 입력된 사용자의 영상은 부가영상대치 과정을 거쳐 부가영상이 대치된 새로운 영상이 얻어진다. 그리고 영상 코덱의 사용권한을 얻은 후, 부가영상이 대치된 영상으로 부호화 과정을 수행한다. 부호화 과정을 통해 얻어진 부호화된 영상 데이터를 네트워크에 연결된 상대편에 전송함으로써 영상 데이터의 입력, 압축, 전송 부분이 정의된다. 네트워크를 통해 수신된 상대방의 영상 데이터는 전송측에서 부호화 된 데이터이므로 복호화를 위해 영상 코덱의 사용권한을 획득하고, 복호화 과정을 수행한다. 복호화된 수신 영상 데이터는 사용자의 화면에 출력된다.

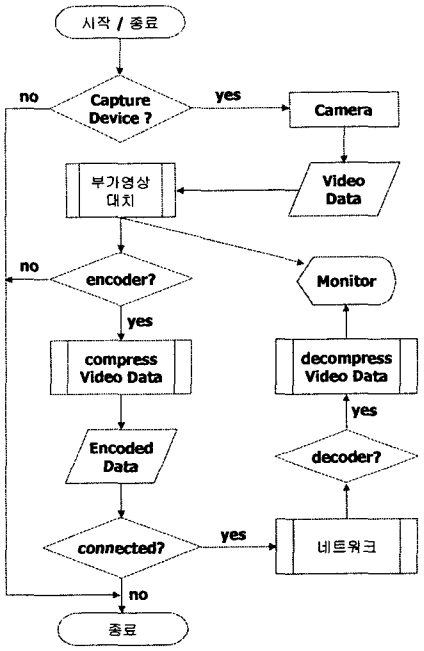

그림 1. 영상 입출력 다이어그램

Fig. 1. Video-Streaming Diagram

음성 입출력은 그림 1에 도시한 영상 입출력과 유사한 과정으로 구현된다. 하지만, 음성 입출력의 경우 부가되는 데이터나 정보가 필요 없으므로, 음성 입출력, 입출력 음성 데이터의 부호화, 복호화 과정에 충실할 수 있게 설계한다. 제안한 프로그램은 실행되면서 음성 입력장치의 유무를 확인한다. 음성 입력장치가 장착되어 있을 경우, 입력장치에 대한 제어권을 획득하고, 사용자의 음성을 입력 받는다. 입력된 사용자의 음성은 음성 코덱의 사용권한이 얻어진 후, 부호화 과정을 거쳐 부호화가 수행된다. 부호화 과정을 통해 얻어진 부호화된 음성 데이터를 네트워크에 연결된 상대편에 전송한다. 네트워크를 통해 수신된 음성 데이터는 복호화를 위해 코덱의 사용권한을 획득하고, 복호화 과정을 수행한다. 복호화된 음성 데이터는 사용자 컴퓨터에 장착된 스피커로 출력된다.

3.2 부가영상대치

영상 입력 장치를 통해 입력된 통화 영상 데이터는 그림 2에 나타낸 순서에 따라 영상 부가영상대치 과정을 통하여 부가영상이 원 영상의 일부분을 대치한 새로운 스트림으로 실시간 부호화되어, 연결된 네트워크를 통해 상대편 컴퓨터로 전송된다.

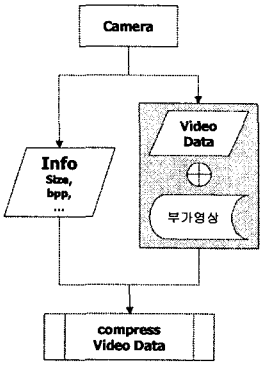

그림 2. 부가영상대치
Fig. 2. Additive Video-Stream Replacement

통화 음성 데이터의 경우, 데이터의 수정 없이 코덱을 통해 부호화되고 전송된다. 수신측에서는, 소켓을 통해 수신된 데이터를 코덱을 통해 복호화하여 출력장치로 전송한다. 전송측에서 전송한 영상데이터는 사용자의 통화 영상에 부가영상이 대치된 영상이므로, 수신측에서는 전송자의 원 영상이 아닌 부가영상이 대치된 영상을 보게 된다.

부가영상대치 과정은 기존의 화상통화 프로그램이나 시스템에서 사용자의 통화 영상만 전달하는 방식과 달리, 통화 중인 사용자의 영상의 일부분을 부가영상으로 대치하여 본 시스템을 사용하는 사용자에게 통화영상과 함께 부가된 정보를 전달할 수 있게 되는 기본 과정으로, 영상 출력 단이나, 음성 입출력과 관계없이 영상 입력 시에만 수행된다.

3.3 입출력 데이터 정의

제안한 시스템에서 입출력되는 데이터가 네트워크를 통해 전송되기 위해선 상대적으로 큰 대역폭을 필요로 한다. 따라서, 입력받은 영상 데이터와 음성 데이터는 압축된 상태로 크기를 줄여서 전송되어져야 하고, 수신측에서는 이를 원 상태로 변환해 출력하는 과정이 요구되어진다. 영상, 음성 데이터를 압축하고 이를 복원하는 과정은 코덱을 이용해 구현할 수 있다.

제안한 시스템을 구현하기 위해, H.263 영상 코덱과, G.723.1 음성 코덱을 사용한다. H.263과, G.723.1은 각각 다양한 화상통화 표준안에서 지원하는 보편적인 영상, 음성 표준이며, 네트워크를 통한 데이트의 송수신에 적합한 크기의 출력 데이터를 생성한다. 또한, H.263과 G.723.1은 ITU-T에서 제안한 화상통화 표준안 중, PSTN에 기반 한 화상통화 표준안인 H.324의 기본 규격이기도 하다. H.263, G.723.1 표준은 각각 다양한 규격의 데이터 형식을 지원하며, 본 논문에서는 통화 영상, 부가영상의 대치 등을 고려해 H.263이 지원하는 규격인 CIF형식에 맞춰 352x288 크기를 갖는 영상 입력을 사용한다. 그리고, G.723.1이 지원하는 규격 중 11,025 Hz, 16 비트, 모노 형식을 갖는 음성 입력을 사용하고, 이에 따라 제안한 시스템을 구현한다.

4. 시스템 구현

구현은 영상, 음성 데이터의 입·출력을 기반으로 한다. 입·출력 데이터의 제어권을 확보한 후, 이로부터 입·출력 데이터 형식을 지정하고, 각각 영상·음성 데이터에 해당하는 코덱을 사용하여, 부호화하고 복호화한다. 부호화/복호화된 데이터는 윈도우 소켓을 사용하여 송수신한다. 구현할 전체 시스템은 아래 그림 3과 같이 구성된다.

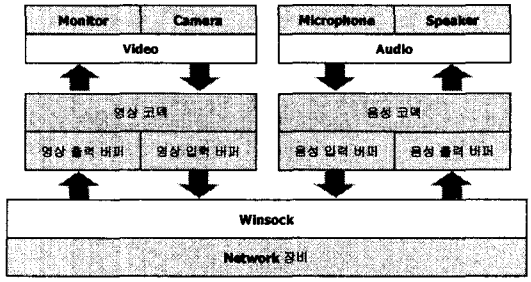

그림 3. 시스템 구성도
Fig. 3. System Diagram

4.1 데이터 입출력

영상 입출력 부분의 구현은 영상 입출력 부분과 입력된 영상을 부호화하고, 출력할 영상을 복호화하는 코덱 부분으로 나누어 구현하였다.

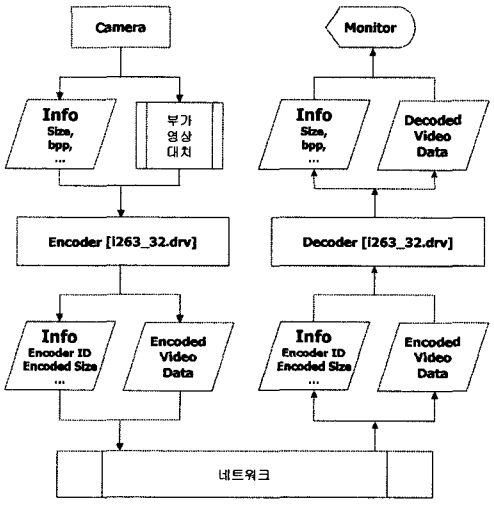

그림 4. 영상 데이터 제어 다이어그램
Fig. 4. Video-Stream Control Diagram

시스템에 장착된 영상 입력 장치를 통해 입력된 사용자의 통화영상에서 통화영상 데이터와 영상 데이터의 형식을 추출한다. 추출된 영상 데이터의 형식과 영상 데이터를 기반으로 코덱을 구동하면, 코덱은 부호화된 영상의 프레임 데이터와 부호화된 데이터의 형식을 반환해준다.

다른 영상 부호화 방식과 유사하게, H.263 규격 또한 부호화된 각 프레임의 크기가 일정하지 않기 때문에, 부호화된 데이터의 크기나 영상 데이터의 가로, 세로 픽셀 수는 네트워크를 통해 수신받은 부호화된 영상 데이터를 복호화할 경우나, 네트워크를 통하여 데이터를 전송할 경우에 사용하기 위해 별도로 저장해 둔다.

4.2 부가영상대치

제안한 시스템은 초당 15 프레임의 통화 영상을 입력 받아 이를 네트워크를 통해 전송한다. 즉, 부가영상의 대치는 매초 15번씩 이루어져야 하기 때문에 부가영상대치를 원활히 구현하기 위해 비트맵 파일로 제작하였다. 부가영상 비트맵은 서로 다른 내용의 두 종류로 제작되었고, 각각 352x48 크기로 제작하였다.

부가영상으로 제작된 비트맵은 352x288 크기의 통화 영상 윗 부분과 아래 부분에 대치되어 그림 5와 같은 새로운 규격을 가지는 영상 데이터가 된다.

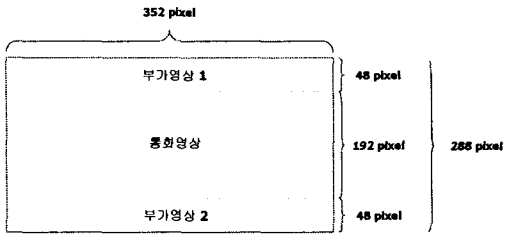

그림 5. 부가영상 구성

Fig. 5. Additive Video-Stream Format

구현에 사용한 부가영상 1은 20개의 비트맵 파일, 부가영상 2는 50개의 비트맵 파일을 사용한다. 각 부가영상은 연속하여 보여질 경우, 동영상을 보는 효과를 낸다. 즉, 각 부가영상을 순차적으로 원 영상에 대치할 경우 통화 영상의 위 부분과 아래 부분에 부가 동영상이 연속하여 재생되는 화면으로 상대편 통화자에게 보여진다.

4.3 전체 구현

구현된 시스템은 입·출력 데이터 형식을 지정하고, 각각 음성·영상 데이터에 맞는 코덱을 사용하여, 부호화/복호화한다. 부호화/복호화된 데이터는 소켓을 사용하여 송수신한다. 입력 데이터는 화상통화에 가장 많이 사용되고 있는 형식으로 지정한다. 영상 입력은 USB 카메라를 사용하였고, 형식은 가로 352픽셀, 세로 288픽셀 초당 15프레임을 가지는 영상 데이터 스트림이다. 음성 입력은 마이크를 통하여 입력된 11,025 Hz, 16 비트, Mono 형식을 가지는 PCM 웨이브이다.

입력된 데이터는 해당하는 코덱에 의해 부호화되어지며, 부호화에 사용한 코덱은, 「Intel I.263 Video Driver v2.55.014」 H.263 코덱과, 「Microsoft G.723.1 Audio CODEC」 G.723.1 코덱이다. 코덱을 통해 부호화된 영상 데이터는 H.263 CIF 형식을 가진다. CIF형식은 ITU-T의 H.263 권고안에서 지원하는 형식 중의 하나로, 입력데이터와 같은 가로 352픽셀, 세로 288픽셀의 크기를 가진다. 부호화된 음성 데이터는, 입력받은 음성 데이터와 같은 형식을 가지는 G.723.1 PCM 웨이브이다.

구현된 프로그램은 서버 모드 또는 클라이언트 모드로 동작할 수 있고, 통화 시 동작은 서버 혹은 클라이언트 중 하나의 모드로 제한한다. 그림 6은 클라이언트 모드로 동작하는 프로그램의 실행화면이며 서버 모드로 동작하는 상대편에 연결하는 방식으로 통화가 성립된다.

그림 6. 클라이언트 모드 동작화면

Fig. 6. Operation Screen in Client Mode

5. 결론 및 향후 연구방향

본 논문에서는 부가영상 정보가 포함된 영상 송수신이 가능한 화상회의 시스템을 구현하였다. 구현된 시스템은 기존의 화상통화 시스템에서 사용자간의 통화 영상만 전달했던 것과 달리 사용자의 통화 영상의 일부분을 부가영상으로 대치하여 화상통화에 참가한 사용자들에게 통화 영상과 함께 부가영상을 전달한다. 제한된 네트워크 환경에서 통화 음성, 통화 영상을 효과적으로 전달하기 위해 음성 데이터, 영상 데이터를 코덱을 사용해 부호화, 복호화하고, 네트워크를 통해 연결된 컴퓨터간에 송수신 데이터를 흐름 제어하였다.

향후 연구과제로는, 동일한 네트워크 대역폭에서, 더 좋은 화질을 제공하거나, 더 큰 화면을 제공하는, 또는 더욱 깨끗한 통화음을 지원하는 코덱이나 표준안이 제정될 경우, 새 시스템에 적용할 수 있도록 유연성을 확보하는 것이다. 또 다른 하나는, 기존의 화상통화 표준안을 따르는 다른 시스템, 또는 장비들과 서로 연동될 수 있도록 호환성을 갖추는 것이다.

참고문헌

[1] 이성우, "H.261 비디오 스트림상의 부가영상 삽입 알고리즘개발," *한국정보과학회 학술발표논문집(B)*, 제27권 1호, 한국정보과학회, 2000.

[2] 정제창역, *그림으로 보는 응용 MPEG*, 교보문고, 1997.

[3] ITU-T Recommendation H.263, Video coding for low bit rate communication.