딥러닝 기반 방송 콘텐츠 클래스 분류 시스템 개발

*김신, **윤경로
건국대학교 컴퓨터공학과
*new.xin22@gmail.com

Development of Broadcast Content Class Classification System based on Deep Learning

*Shin, Kim, **Kyoungro, Yoon
Department of Computer Science Engineering, Konkuk University

요 약

최근 수 년간 비디오 콘텐츠 소비 공간이 인터넷으로 확장되며 지능적 비디오 콘텐츠 추천 기술 개발이 진행되어 왔다. 하지만 지능적 비디오 콘텐츠 추천 기술은 사용자의 기호나 업로드된 비디오 콘텐츠의 제목 등을 기반으로 하여 비디오 콘텐츠 클래스에 대한 분석 없이 유사한 비디오 콘텐츠를 탐색하고 추천해주는 기술이 대부분이다.

본 논문에서는 지능적 콘텐츠 추천을 위한 딥러닝 기반 방송 콘텐츠 클래스 분류 시스템을 제안한다. 방송 콘텐츠 내 영상 정보를 이용하여 방송 콘텐츠 클래스를 분류하며 높은 분류 정확도를 보여주는 것을 확인할 수 있다.

1. 서론

최근 수 년간 비디오 콘텐츠를 소비할 수 있는 공간이 인터넷 상으로 확장되면서 방송 콘텐츠 뿐만 아니라 인터넷 방송 콘텐츠도 상일적으로 소비할 수 있게 되었다. YouTube 와 같은 무료 동영상 공유 사이트에서 사용자가 특정 단어를 검색하여 비디오 콘텐츠를 검색할 때, 해당 비디오 콘텐츠와 연관된 추천 목록을 확인할 수 있다. 하지만 추천 비디오 콘텐츠는 사용자의 비디오 콘텐츠 선호도나 업로드된 비디오 콘텐츠 제목 등으로 구성된다. 즉, 재생되고 있는 비디오 콘텐츠에 대한 분석없이 추천 목록이 형성된다.

비디오 콘텐츠 클래스에 대한 분석 없이 비디오 콘텐츠 추천 목록이 형성되기 때문에 해당 콘텐츠 클래스와 연관 없는 추천 콘텐츠들이 포함되기도 한다. 따라서 지능적 비디오 콘텐츠 추천 시스템의 정확도를 올리기 위해서는 해당 비디오 콘텐츠 분석도 요구된다.

본 논문에서는 딥러닝 기반 방송 콘텐츠 분류 시스템을 제안한다. 현재까지는 방송 콘텐츠가 비디오 콘텐츠 중 가장 다양한 소비층을 가지고 있으며 파급력이 강력하기 때문이다.

본 논문의 구성은 다음과 같다. 2 절에서는 방송 콘텐츠 클래스를 분류에 필요한 배경 지식에 대해 설명하며, 3 절에서는 방송 콘텐츠 클래스 분류 시스템을 제안한다. 4 절에서는 본 논문에서 제안하는 방송 콘텐츠 클래스 분류 시스템에 대한 실험 결과에 대해 서술하며 마지막으로 5 절에서는 본 논문의 결론을 맺는다.

2. 배경 지식

방송 콘텐츠 클래스 분류 시스템은 이미지 분류 기술을 필요로 한다. 이미지 분류 기술은 최근 딥러닝 기법 도입을 통해 엄청난 분류 정확도 향상을 가져왔다.

GoogLeNet[1]은 2014 년 ILSVRC(ImageNet Large Scale Visual Recognition Challenge)에서 1 등을 차지한 이미지 분류 모델로 통상적으로 Inception v1 으로 불린다. Inception 모듈은 CNN 망의 깊이를 깊게 만들어주되 연산량은 크게 증가시키지 않는 것이 가장 큰 특징이며 현재 Inception v3[2]가 일반적으로 가장 많이 사용되고 있는 이미지 분류 모델 중 하나이다.

ResNet[3]은 2015 년도 ILSVRC 에서 우승한 모델로 계층을 일부 뛰어넘는 기술인 Residual Connection 의 도입으로 계산량은 낮추면서 높은 정확도의 이미지 분류를 가능케 했다.

Inception-ResNet[4]은 2016 년 ILSVRC 에서 우승한 이미지 분류 모델로서 Inception 모듈과 Residual Connection 을 융합하여 보다 높은 이미지 분류 정확도를 보여준다. 세 모델의 이미지 분류 에러율 비교는 표 1 을 통해 확인할 수 있다.

표 1. 네트워크에 따른 이미지 분류 오류율 비교(퍼센트)[4]

Network	Resnet[3]	Inception v3[2]	Inception-Resnet(v2)[4]
Top-5 Error(%)	7.8%	4.6%	4.1%

3. 방송 콘텐츠 클래스 분류

본 논문에서는 딥러닝 기반 방송 콘텐츠 클래스 분류 시스템을 제안한다. 방송 콘텐츠 클래스 분류를 위한 딥러닝 모델은 2016 년 ILSVRC 우승 모델인 Inception-Resnet[4]를 차용하였으며 transfer learning 을 통해 방송 콘텐츠 클래스 분류 모델을 학습하였다.

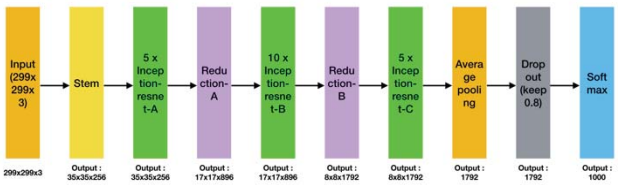

그림 1. Inception-Resnet 모델 기본 구조[4]

4. 실험 결과

본 논문에서는 실험을 위해 News, Entertainment, Documentary, Drama, Concert, Sport 로 구성된 6 개의 방송 콘텐츠 클래스를 정의하고 클래스에 해당하는 데이터 세트를 자체 구축하였으며 방송 콘텐츠 클래스 분류 정확도를 측정하였다. 방송 콘텐츠 클래스 분류 모델 훈련 및 실험 환경은 표 2 과 같다.

표 2. 방송 콘텐츠 클래스 분류 모델 훈련 및 실험 환경

Operating System	Ubuntu 16.04 LTS
GPU	GeForce GTX 980 ti
Framework	Keras, Tensorflow
Language	Python 3.6

그림 2. 방송 콘텐츠 클래스 분류를 위한 훈련 데이터 예시

방송 콘텐츠 클래스 분류 모델을 위한 데이터 세트는 각 방송 콘텐츠 클래스당 훈련용 이미지 940 장, validation 용 이미지 260 장, 분류 실험용 이미지 360 장으로 구성되어 있다. 스케일 조정, 이미지 좌우 상하 반전, 이미지 회전을 통해 훈련용 이미지 데이터를 증대시켰으며 그림 2 는 방송 콘텐츠 클래스 분류 모델을 훈련하기 위한 데이터의 일부 예제이다.

표 3 를 통해 방송 콘텐츠 클래스 분류 실험의 정확도를 확인할 수 있으며 실험 결과 Documentary 클래스를 제외한 모든 방송 콘텐츠에서 약 90%의 정확도로 클래스 분류가 가능한 것을 확인할 수 있다.

표 3. 방송 콘텐츠 클래스 분류 정확도(퍼센트)

Class	News	Enter-tain-ment	Docu-men-tary	Drama	Con-cert	Sport
Classification Precision(%)	80.3	93.1	55.8	89.7	87.8	97.8

5. 결론 및 향후 과제

본 논문에서는 딥러닝 기반의 방송 콘텐츠 클래스 분류 시스템을 제시하였다. 이미지 분류 알고리즘을 이용해 방송 콘텐츠 클래스를 분류하였으며 높은 분류 정확도를 보여주는 것을 확인할 수 있었다.

향후 연구 방향으로 방송 콘텐츠 클래스 분류를 통해 얻은 비디오 클래스 기반 지능적 비디오 콘텐츠 추천 엔진 및 시스템을 연구해야 할 것이다.

감사의 글

본 논문은 2018 년 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(2017-0-00024, UHD 방송콘텐츠 기반 지능형 Dynamic Media 생성, 분배 및 소비 기술 개발)

참 고 문 헌

[1] Szegedy, Christian, et al. "Going deeper with convolutions." Cvpr, 2015.

[2] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

[3] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

[4] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." *AAAI*. Vol. 4. 2017.

검색어 생성을 위한 딥 러닝 기반 문장 분석 연구

*나성원 **윤경로

건국대학교

*securityin4@naver.com

Deep Learning based Sentence Analysis for Query Generation

*Seong-Won, Na **Kyoungro, Yoon

Konkuk University

요약

최근 이미지의 Visual 정보를 추출하고 Multi label 분류를 통해 나온 결과의 상관관계를 modeling하여 문장으로 출력하는 CNN-RNN 아키텍처가 많은 발전을 이뤘다. 이 아키텍처의 출력은 이미지의 정보가 요약되어 문장으로 표현되기 때문에 Semantic정보가 풍부하여 유사 콘텐츠 검색에도 사용 가능하다. 하지만 결과 문장에 사람이 포함 되면 광범위한 검색 결과를 얻게 되고 부정확한 결과를 초래하게 된다.

이에 본 논문에서는 문장에서 사람을 인식하여 Identity를 부여함으로써 검색어를 좀 더 구체적으로 생성하고자 한다. 이 문제를 해결하기 위해 자연어 처리의 분야 중 하나인 개체명 인식(Named Entity Recognition) 문제로 다루며, 가장 많이 사용되고 있는 모델인 Bidirectional-LSTM-CRF와 CoNLL2003 dataset을 사용하여 수행 한다.

1. 서론

컴퓨터 비전 분야에서 이미지를 단순히 인식, 분류하는 문제는 ILSVRC15 dataset에 대해 Top5 error rate 3.08%를 달성하게 되면서 사람을 능가하는 성능에 도달하였다. 그에 따라 자연스럽게 단순 인식, 분류 문제를 다루는 것이 아닌 이미지내의 모든 object, properties, attributes, action 등을 처리하는 Multi-label classification, tagging, captioning과 같은 이미지 속성에 대해 보다 풍부한 설명을 생성하는 문제에 관심이 증가하였고, 많은 발전이 이루어 졌다[1]. 이 같은 문제를 다룰 때 가장 많이 사용되는 기술은 CNN-RNN구조로 CNN은 이미지의 feature를 추출해 RNN의 입력으로 사용하고, RNN은 concept prediction과 label-correlation modeling을 수행하여 정렬 된 문장을 생성하는 구조이다. 이 결과로 나온 문장은 이미지의 semantic한 정보를 담고 있기 때문에 유사 콘텐츠를 검색 시 검색어로 사용 가능하다. 예를 들어 “a man is playing tennis on a tennis court”와 같은 문장이 결과로 출력되면 특정한 선수가 아닌 남성이 테니스를 하고 있는 비교적 큰 category를 의미하는 검색어가 될 것이다. 위와 같은 문장을 보다 구체적인 검색어로 사용하기 위해 문장에서 사람을 인식하고, identity를 추가적으로 포함한 검색어를 생성하고자 한다. 여기에서 identity라고 하면 위 문장에서 ‘man’이라는 단어를 인식하여, ‘정현’이라는 고유 명사로 대체하는 것을 의미 한다.

자연어 처리의 task 중 하나인 개체명인식(Named Entity Recognition)은 문장에서 나타나는 고유한 의미를 가지는 명사를 인식하는 것으로 주로 4Class인 인명(Person), 지명(Location), 기관명(Organization), 기타(MISC)로 나눌 수 있다[2]. 이 NER 문제를 처리하는데 많이 사용되는 모델 중 하나인 Bidirectional-LSTM-CRF모델은 여러 분야에서 각광 받고 있는 Long Short-term Memory Network(LSTM)[3]기반 RNN이며, 기존의 문제였던 gradient vanishing문제를 해결하기 위해 제안 되었고, Sequence한 문제에서 강력한 성능을 보여 언어 모델, 음성 인식, 자연어 이해등과 같은 분야에서 많이 사용 되고 있다. 문장 분석은 위에서 설명한 모델을 사용하고, dataset으로는 CoNLL2003을 사용하여 학습 하였다.

2. 본론

2.1 Dataset

CoNLL2003의 NER dataset은 문장 단위로 구성 되어 있고, 문장을 구성하는 단어와 label이 쌍으로 이루어져 있으며, Network로 입력 시 하나의 쌍이 입력되는 구조이다. 하지만 우리가 원하는 사람을 지칭하는 Woman, Man, Player, Singer등과 같은 단어는 PER이라는 label이 달려 있지 않기 때문에 필요한 단어의 label을 직접 수정 하여 훈련을 진행 하였다. 우리 시스템의 목적은 사람을 인식하는 것이기 때문에 다른 label은 수정하지 않고 사용 되었다.

2.2 Bidirectional LSTM CRF 모델

RNN은 sequence data를 처리하는데 사용하는 Neural Network 이며, 이론적으로는 긴 의존성을 배울 수 있지만 실제로는 가장 최근 입력에 편향되는 경향이 있다. Long Short-Term Memory Network는 메모리 셀을 통하여 이 문제를 해결하기 위해 설계 되었으며 장거리 종속성을 포착가능 하다. RNN의 gradient vanishing 문제를 해결한 LSTM RNN은 다음과 같이 정의 된다.

$$i_t = \sigma(W_{xt}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i$$
$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f$$
$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$
$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o)$$
$$h_t = o_t \tanh(c_t)$$
$$y_t = g(W_{hy}h_t + b_y)$$

위 식에서 σ는 sigmoid 함수이고, I, f, o, c는 각각 input gate, forget gate, output gate, memory cell vector를 나타내며 각 벡터의 크기는 hidden layer 벡터 크기와 같다.

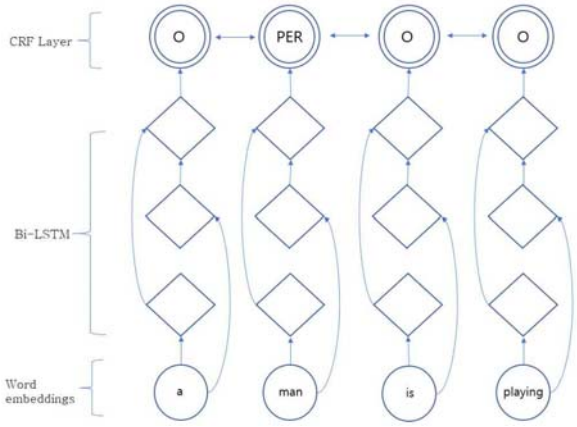

그림 1. Bidirectional LSTM CRF 구조[4]

그림 1은 Bidirectional LSTM CRF모델의 구조를 나타낸다. 기존 LSTM CRF와 달리 양방향으로 학습되기 때문에 현재 label 결정에 이전단어와 다음단어의 정보를 모두 볼 수 있다. Bidirectional-LSTM CRF 모델의 학습을 위해 Stochastic Gradient Descent(SGD)[5]알고리즘을 사용한다.

3. 실험

본 논문에서는 Bidirectional-LSTM-CRF모델에 사용한 영어 word embedding은 Collobert & Westone의 단어 표현을 사용하였고, feature embedding은 평균 0, 분산 0.01이 되도록 무작위로 초기화 시킨 값을 사용 하였다. 과적합 문제를 완화하기 위해 dropout rate는 0.5로 적용하였고, LSTM hidden layer의 dimension은 100으로 설정하여 진행 하였다. 문장에서 사람을 인식하는 것이 목적이기 때문에 dataset을 수정해서 training을 진행 하였고, 평가를 위해 CoLL2003 NER test dataset이 아닌 자체 dataset을 통해 평가 하였다. 이 dataset은 인터넷상에서 수집하였으며, 평가할 문장들은 사람이 포함 된 이미지를 CNN-RNN 아키텍처에 입력으로 사용하여, 처리 후 나온 출력 문장을 사용하였다.

표 1은 training dataset의 tag 변경 전, 후에 대해 학습한 model의 출력 결과를 보여준다. 사람을 표현하는 단어 전체가 아닌 특정단어에 대해 training set을 수정하였기 때문에 93% 이상의 높은 precision이 측정 되었다.

표 1. tag 변경 전과 후의 결과 비교

time	word	변경 전 tag	변경 후 tag
1	A	O	O
2	Man	O	B-PER
3	is	O	O
4	playing	O	O
5	Tennis	O	O
6	on	O	O
7	a	O	O
8	Tennis	B-LOC	B-LOC
9	Court	I-LOC	I-LOC

4. 결론

우리는 CNN-RNN 아키텍처의 출력인 Semantic한 문장을 좀 더 구체적인 query로 생성하기 위해 Bidirectional-LSTM-CRF 모델을 사용하여 문장내의 사람을 인식하는 실험을 진행 하였다. 특정 단어에 대해서만 label을 수정하였기 때문에 정확도는 높게 나왔지만 dataset을 더 크게 구성할 수 있다면 NER task에서 좀 더 넓은 범위의 person 인식을 할 수 있을 것으로 판단된다. 하지만 기존 Bidirectional LSTM CRF모델을 사용하여 실험하였기 때문에 기존에 person으로 잘못 인식한 단어들까지 identity를 갖게 되는 단점을 발견 하였다. 예를 들면 “a Woman is holding a Teddy Bear.”란 문장에서 ‘Woman’과 ‘Teddy Bear’ 두 단어가 PER로 인식 되면서 두 단어에 identity가 들어가게 되어 “IU is holding a IU”란 문장이 결과로 도출 되었다. 이러한 문제점을 해결하기 위해 추후 연구에서는 문장에서 하나의 술어 당 하나의 person만 인식하는 방법을 연구할 계획이다.

감사의 글

이 논문은 2018년 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (UHD 방송콘텐츠 기반 지능형 Dynamic Media 생성, 분배 및 소비 기술 개발)

참고문헌

[1] Liu, Feng, et al. "Semantic Regularisation for Recurrent Image Annotation." arXiv preprint arXiv:1611.05490 (2016).

[2] Huang, Zhiheng, Wei Xu, and Kai Yu. "Bidirectional LSTM-CRF models for sequence tagging." arXiv preprint arXiv:1508.01991 (2015).

[3] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[4] Lample, Guillaume, et al. "Neural architectures for named entity recognition." arXiv preprint arXiv:1603.01360 (2016).

[5] Bottou, Léon. "Large-scale machine learning with stochastic gradient descent." Proceedings of COMPSTAT'2010. Physica-Verlag HD, 2010. 177-186.

MMT 기반의 실시간 대용량 미디어 전달을 위한 병렬 전송 효율 개선

안은빈, 김아영, 원광은, [1]윤재관, 서광덕
연세대학교 컴퓨터정보통신공학부, [1]한국전자통신연구원
09eunbin_ahn@yonsei.ac.kr

Improvement of Parallel Transmission Throughput for Transporting Real-time Mass Media Based on MMT

Eun-bin An, A-young Kim, Kwang-eun Won, [1]Jae-kwan Yoon, Kwang-deok Seo
Division of Computer and Telecommunications Engineering, Yonsei University, Korea
[1]Electronics and Telecommunications Research Institute, Daejeon, Korea

요 약

최근 실시간 대용량 미디어에 대한 사용자의 요구가 증가함에 따라 자연스러운 영상 재생을 위한 전송 기법이 활발히 연구되고 있다. MPEG MMT 는 이러한 차세대 대용량 미디어 전송 규격으로 주목 받고 있다. 하지만 실시간 대용량 미디어의 크기는 점차 커지고 있고 이에 따라 보다 효율적이고 빠른 전송을 위해서 다각도의 연구가 필요하다. 본 논문에서는 MMT 기반의 실시간 대용량 미디어 전송의 개선을 위하여 병렬 전송을 제안하고 이에 따른 MMT 의 활용 방법을 제시한 인터페이스를 소개한다.

1. 서론

최근 VR(Virtual reality), UHD 4K 또는 8K 고화질 영상 등 대용량 미디어 콘텐츠의 사용자 요구가 증가하면서 효율적인 대용량 미디어 전송 기법이 활발히 연구되고 있다. 이에 따라, MPEG 에서는 2014 년 6 월, 이 기종 망에서의 대용량 멀티미디어를 전송하는 기술인 MMT(MPEG-H Part 1: MPEG media transport ISO/IEC 23008-1)를 발표하였다. ATSC 3.0 기반의 핵심 기술인 MMT 는 MPEG-2 TS 의 장점을 최대한 유지하면서 IP 망에 적합하도록 설계되었으며, 고정 패킷 사이즈를 사용하는 MPEG-2 TS 와 달리 IP 망에 적응적으로 사용될 수 있다. 특히 대용량 콘텐츠 전송을 위한 하이브리드 방송에서 MMT 는 실시간 대용량 미디어 전송에 보다 효율적이다. 하지만 MMT 기반의 실시간 대용량 미디어 전송을 실제로 단일 전송 회로로 구성하였을 때 네트워크 망의 상태에 따라 실시간으로 영상을 처리하지 못하는 한계에 부딪힐 수 있다. 실시간 대용량 미디어를 안정적으로 전송하기 위해서는 프레임 단위의 병렬 전송 기법과 이에 따른 MMT 프로토콜의 활용법에 대한 연구가 필요하다.

따라서 본 논문에서는 실시간 대용량 미디어를 MMT 기반으로 전송할 때 보다 효율적으로 전송할 수 있는 병렬 전송 인터페이스를 제안하고 MMT 프로토콜을 병렬 전송에 적용하는 방법을 제안하고자 한다.

2. 본론

2.1 MMT 데이터 모델

MMT 서비스에서 독립적으로 완벽하게 소비가 가능한 단위를 MPU(Media Processing Unit)로 정의한다. MPU 는 timed media 뿐만 아니라 non-timed media 를 포함할 수 있고, timed media 의 경우 MPU 에는 하나 이상의 AU (access unit)가 포함된다. 하나의 AU 를 분할하여 여러 개의 MPU 에 나누어 담는 것은 금지되어 있다. 또한 MPU 는 ISOBMFF 와 호환 가능한 파일포맷으로 인캡슐레이션이 가능하다.

Asset 은 동일한 Asset ID 를 갖는 하나 이상의 MPU 를 묶어서 만들지게 되며 표현 정보 (PI: presentation information) 및 전송 특성 (TC: transport characteristics)이 부여되는 가장 큰 단위가 된다. Asset ID 를 통해서 MPU 가 소속된 Asset 을 파악할 수 있다.

Package 하나 이상의 Asset 들과 이 Asset 에 연관된 PI 및 TC 를 모두 포함하는 논리적인 데이터 구조이다. 그림 1 은 MPU 와 Asset, TC, PI 를 포함한 Package 의 구조를 보여준다.

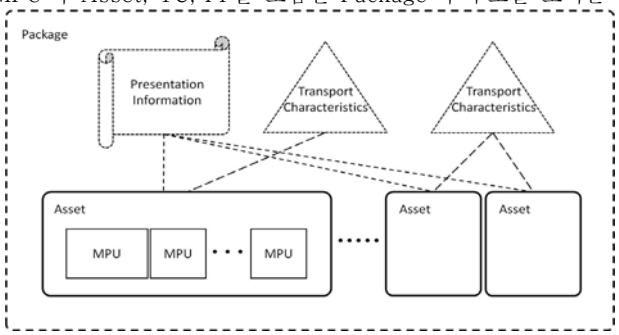

그림 1 MMT Package 의 구조

2.2 MMT 기반의 대용량 미디어 병렬 전송 인터페이스

본 논문에서는 다중 포트를 구성하여 병렬 전송을 구현하였다. 그림 2 는 MMT 기반의 병렬 전송 인터페이스를 보여주며 이때 전송되는 미디어 스트림은 Video 와 Audio 로

한정한다.

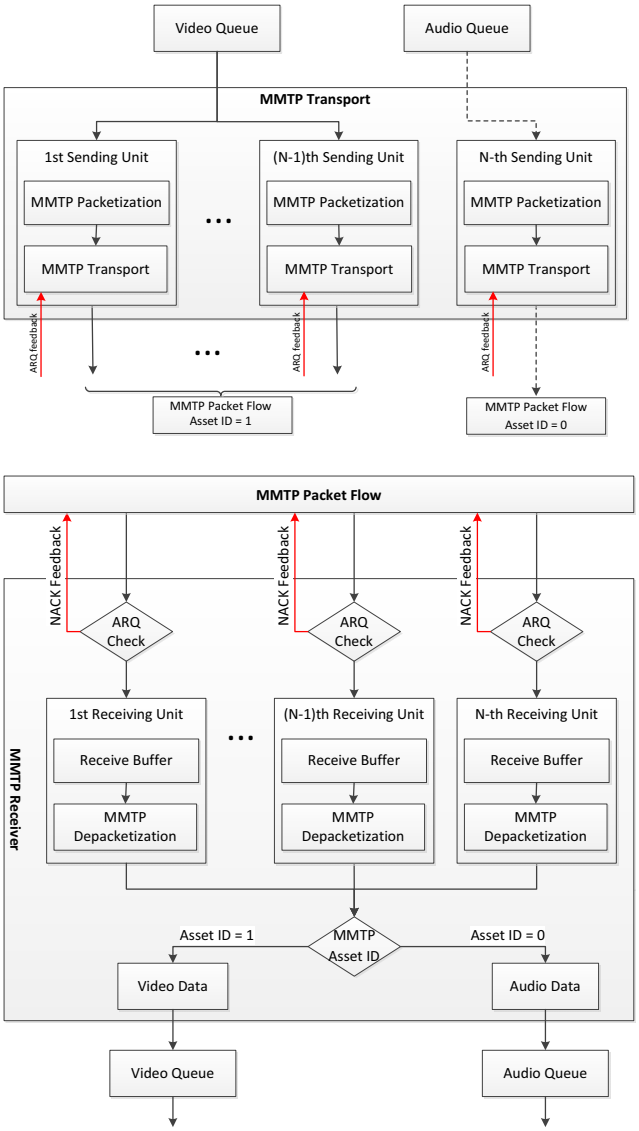

그림 2 MMT 기반의 병렬 전송 인터페이스

장치로부터 캡처되고, 인코딩된 Video 와 Audio 스트림 데이터는 Video Queue 와 Audio Queue 에 각각 저장된다. Queue 에 쌓여진 데이터들은 MMTP Transport 에서 MMTP Packetize 되고 전송된다. Audio 의 데이터는 상대적으로 크기가 작기 때문에 하나의 Sending Unit 을 사용하지만 Video 데이터들은 1 부터 N-1 까지의 병렬로 구성된 채널로 전송된다. 각 Sending Unit 은 영상의 한 프레임을 전담하여 전송하며, 한 프레임의 전송을 마치면 다른 Sending Unit 들이 현재 전송하고 있지 않은 다음 프레임을 전송하게 된다. 이때 모든 Sending Unit 들은 다른 Sending Unit 들과 완벽하게 독립적으로 데이터를 전송한다.

전송 받은 MMTP Packet 들은 MMTP Receiver 에서 ARQ(Retransmission Error Control) 모듈을 거쳐 Sending Unit 과 정확하게 쌍을 이루는 Receiving Unit 에 데이터를 전달한다. Receiving Unit 의 Receive Buffer 는 한 프레임을 구성하기 위한 공간으로 한 프레임의 모든 데이터가 수신되면 다시 한 프레임으로 재구성하여 Video Queue 와 Audio Queue 에 각각 쌓이고 랜더링한다.

ARQ 모듈은 병렬 전송에 최적화하여 각 Receiving Unit 앞에 붙이고 MMTP Receiver 에서 수신이 올바르지 않은 경우에 NACK 신호를 서버에 전송한다. MMTP Transport 는 NACK 신호를 수신했을 경우에 새로운 데이터를 전송하지 않고 직전에 보냈던 데이터를 재전송한다. 오직 Receiver 에서만 에러를 확인하는 비대칭 ARQ 모듈로서 병렬 전송의 효율을 높이고 또한 실시간 미디어에서 ARQ 적용으로 인하여 딜레이가 발생하지 않게 하기 위해 고안한 최적화 모듈이다.

2.3 병렬 전송에서 MMT 의 활용

대용량 미디어를 MMT 기반으로 전송하면 완전하고 독립적으로 처리될 수 있는 부호화된 미디어 데이터 유닛을 의미하는 MPU 를 통신망으로 한 번에 보내질 수 있는 전송 유닛의 크기인 MTU(Maximum Transport Unit) 크기에 맞춰 나누어 전송해야 한다. 일 예로 MPU 는 1 초의 비디오를 구성하는 1 GOP 로 구성될 수 있고, MFU(Media Fragment Unit)는 각 픽쳐 프레임을 포함할 수 있다. 하지만 대용량 영상(4K 영상 이상)의 압축된 한 프레임의 크기가 MTU 보다 큰 경우가 대부분이기 때문에 MMTP Packet 은 잘게 쪼개진 프레임의 일부를 보내게 된다. 따라서 잘게 쪼개진 프레임을 합성하는 과정이 필요하며 병렬로 구성된 각 채널이 한 프레임씩 전송하여 다시 한 프레임을 재구성 할 수 있도록 구현하였다. 이러한 방식은 단위 Sending Unit 이 각 픽쳐 프레임 별로 처리하게 함으로써 병렬 전송으로 인하여 생길 수 있는 데이터의 비순차 수신의 빈도를 줄일 수 있고, Ordering 에 들이는 시간을 줄여 실시간 전송에 더 효율적이다. 또한, Receive Buffer 의 설정에 따라 픽쳐 프레임, GOP 등의 단위로 받을 수 있기 때문에 전송 상황에 따른 차별화된 전송이 용이하다

5. 결론

본 논문에서는 실시간 대용량 미디어를 MMT 기반의 병렬 전송하는 인터페이스를 제안하고 이때 MMT 프로토콜의 활용 방법을 분석하였다. MMT 는 실시간성이 뛰어나고 대용량 미디어 전송에 적합한 프로토콜이나 앞으로 발전하는 미디어 시장에 맞추기 위해서는 전송 효율을 보다 높일 수 있는 추가적인 기술 연구가 필요하다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 '범부처 Giga KOREA 사업[GK17P0100, Giga Media 기반 Tele-Experience 서비스 SW 플랫폼 기술 개발]'의 지원을 받아 수행하였음"

참고문헌

[1] ISO/IEC 23008-1, High efficiency coding and media delivery in heterogeneous environments - MPEG-H Part 1: MPEG Media Transport (MMT), 2014.

[2] Jung, Tae-Jun, Hong-rae Lee, and Kwang-deok Seo. "Overview on MPEG MMT Technology and Its Application to Hybrid Media Delivery over Heterogeneous Networks." Pacific Rim Conference on Multimedia. Springer, 2015.

[3] 손예진, "이 기종 망에서의 UHD 비디오 전송을 위한 MMT 기반 방송시스템 설계", 방송공학회논문지, 제 20 권 제 1 호, 2015 년 1 월.

[4] Kim, Chang-Ki, et al. "Apparatus and method for configuring mmt payload header." U.S. Patent Application No. 15/031,540.

HTTP 적응적 스트리밍에서 UHD 콘텐츠의 효율적인 대역폭 활용을 위한 세그먼트 전송 기법

김희광, 정광수
광운대학교 전자통신공학과
hkkim@cclab.kw.ac.kr, kchung@kw.ac.kr

Segment Scheduling Scheme for Efficient Bandwidth Utilization of UHD Contents Streaming in HTTP Adaptive Streaming

Heekwang Kim, Kwangsue Chung
Department of Electronics and Communications Engineering, Kwangwoon University

요 약

최근 네트워크 기술과 스마트 단말의 보급으로 인해 비디오 스트리밍 서비스에 대한 수요가 증가하게 되었다. 네트워크를 효율적으로 사용하여 비디오 스트리밍 서비스를 제공하기 위해 적응적으로 전송률을 조절하는 HTTP (HyperText Transfer Protocol) 적응적 스트리밍 서비스가 주목 받게 되었다. UHD (Ultra High Definition) 콘텐츠는 HD (High Definition) 콘텐츠에 비해 적어도 4 배 이상의 크기를 갖기 때문에 끊김 없는 UHD 콘텐츠 스트리밍 서비스를 제공하기 위해서는 많은 가용 대역폭이 필요하다. 기존의 HTTP 적응적 스트리밍 방식은 정상 상태 (Steady State)에서 가용 대역폭보다 낮은 품질의 비디오 세그먼트를 일정 시간마다 주기적으로 요청하여 다운로드 받는다. 정상 상태에서는 가용 대역폭과 콘텐츠의 인코딩 율에 차이에 따라 On-Off 구간의 패턴이 반복되어 발생하고, 빈번한 Off 구간에 의해서 대역폭이 낭비되는 문제점이 있다. 따라서 본 논문에서는 HTTP 적응적 스트리밍에서 UHD 콘텐츠의 효율적인 대역폭 활용을 위한 세그먼트 전송 기법을 제안한다. 제안하는 기법은 Off 구간의 빈도수를 줄이기 위한 집단 세그먼트 전송 방식과 대역폭 낭비를 최소화 하기 위한 세그먼트 품질 조절기법으로 구성되어 있다.

1. 서론

네트워크 기술과 스마트 단말의 보급으로 인해 비디오 스트리밍 서비스에 대한 수요가 증가하게 되었다. 이에 따라 네트워크 상태 변화에 적응적으로 전송률을 조절하여 사용자 체감 품질을 향상시키는 HTTP (HyperText Transfer Protocol) 적응적 스트리밍 서비스가 주목 받게 되었다. HTTP 적응적 스트리밍 서비스는 서버에 다양한 비트율로 인코딩 된 세그먼트로 구성되어 있고, 클라이언트의 요청에 따라 세그먼트를 전송하는 방식이다. 이 때 클라이언트는 네트워크 가용 대역폭을 측정하고 다음에 요청할 품질을 선택하여 세그먼트를 요청한다 [1].

최근 Youtube 나 Netflix 같은 멀티미디어 스트리밍 업체들은 UHD (Ultra High Definition) 콘텐츠 스트리밍 서비스를 제공하고 있다. UHD 콘텐츠의 특징은 ITU-R BT. 2020 과 SMTPE ST 2036-1 에 명시되어 있다 [2]. UHD 콘텐츠는 데이터의 크기를 줄이기 위해 기존의 HD (High Definition) 콘텐츠에 비해 GoP (Group of Picture)의 길이가 긴 특징을 가지고 있지만, 여전히 HD 콘텐츠에 비해 적어도 4 배 이상의 크기를 갖는다. 따라서 끊김 없는 비디오 스트리밍 서비스를 제공하기 위해서는 HD 품질에 비해 많은 가용 대역폭이 필요하다.

HTTP 적응적 스트리밍 방식은 네트워크 변화에 반응하기 위해서 클라이언트가 측정한 가용 대역폭 보다 낮은 품질의 비디오 세그먼트를 지속적으로 요청하여 다운로드 받는다. 기존의 세그먼트 전송 기법은 초기에 대역폭 낭비를 방지하고, 안정적인 버퍼의 양을 유지하기 위해 연속적으로 세그먼트를 요청하는 버퍼링 상태 (Buffering State)로 동작한다. 버퍼가 안정적인 상태가 되면 버퍼의 오버플로우를 방지하기 위해 주기적으로 세그먼트를 요청하는 정상 상태 (Steady State)로 동작한다. 정상 상태에서 가용 대역폭과 콘텐츠의 인코딩 율에 차이에 따라 다운로드 받는 구간 (On)과 버퍼를 소비하는 구간 (Off)으로 On-Off 패턴이 반복되어 발생한다 [3]. Off 구간은 데이터가 전송되지 않는 구간을 나타내며, 낭비된 대역폭을 의미한다. 기존의 세그먼트 전송 방식의 정상 상태에서 콘텐츠를 전송할 경우 서버에 구성된 콘텐츠의 품질과 클라이언트의 품질 조절 기법에 의해 주기적으로 Off 구간이 발생하고 대역폭이 낭비된다. UHD 콘텐츠의 경우 다른 품질의 콘텐츠에 비해 높은 인코딩 율을 갖고 있기 때문에 가용 대역폭에 따라 대역폭 낭비가 심화되는 문제점을 가지고 있다.

본 논문에서는 HTTP 적응적 스트리밍에서 UHD 콘텐츠의 효율적인 대역폭 활용을 위한 세그먼트 전송 기법을 제안한다. 제안하는 기법은 집단 세그먼트 전송 방식과 집단 세그먼트 품질 조절 기법으로 구성되어 있다. 집단 세그먼트 전송 방식은 한번의 요청 메시지에 여러 세그먼트를 동시에 요청하여 다운로드 받는다. 여러 세그먼트를 연속적으로 받음으로서 Off 구간의 발생 빈도를 감소시켜 낭비된 대역폭을 줄인다. 집단 세그먼트 품질 조절 기법은 측정된 가용 대역폭과 버퍼 상태를 이용

하여 한번의 요청 메시지에 요청할 세그먼트의 수와 품질을 결정한다. 요청 세그먼트의 수와 품질은 효율적인 전송을 위해 낭비된 대역폭이 가장 적은 품질을 결정한다.

2. 제안하는 세그먼트 전송 기법

2.1 집단 세그먼트 전송 기법

기존 세그먼트 전송 기법은 그림 1 과 같다. 전송 초기에 안정적인 버퍼 상태를 위해 버퍼링 상태로 동작한다. 버퍼링 상태는 버퍼를 안정적으로 관리하기 위해 빠르게 버퍼를 채우는 단계이다. 따라서 하나의 세그먼트가 다운로드 되면 바로 다음 세그먼트를 요청하여 빠르게 버퍼를 채운다. 버퍼가 일정 임계값에 도달하게 되면 클라이언트는 버퍼 오버플로우를 방지하기 위해 정상 상태로 동작하게 된다. 정상 상태에서는 주기적으로 세그먼트를 요청하고, 기존 기법에서는 세그먼트의 길이 마다 다음 세그먼트를 요청한다. 주기적으로 세그먼트를 요청하기 때문에 다운로드 받은 콘텐츠의 품질과 가용 대역폭의 차이에 따라 데이터를 전송하지 않는 Off 구간이 빈번하게 발생한다. Off 구간은 가용 대역폭의 낭비를 의미하며 콘텐츠의 품질과 가용 대역폭의 차이가 클수록 많은 대역폭 낭비가 발생한다.

제안하는 집단 세그먼트 전송 기법은 그림 2 와 같다. 전송 초기에는 기존 기법과 같이 버퍼링 상태로 동작한다. 버퍼가 임계값에 도달하게 되면 클라이언트는 집단 세그먼트 전송 상태로 동작한다. 집단 세그먼트 전송 상태는 요청주기 T초 마다 세그먼트를 집단 단위로 요청한다. 요청 세그먼트 집단의 수와 품질은 가용 대역폭 예측값과 버퍼 상태를 기반으로 대역폭 낭비가 최소화 되는 값을 결정한다. 다수의 세그먼트를 집단 형식으로 요청 함으로서 Off 구간의 빈도를 줄이고, 대역폭 낭비가 최소화 되는 세그먼트의 수와 품질을 결정함으로써 기존 기법에 비해 높은 품질을 요청 할 수 있다.

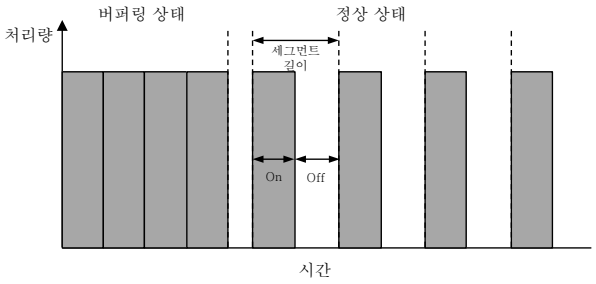

그림 1. 기존 세그먼트 전송 기법

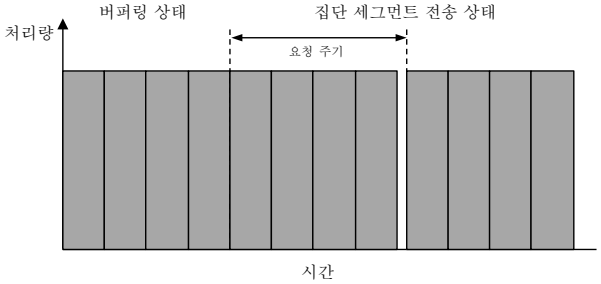

그림 2. 집단 세그먼트 전송 기법

2.2 집단 세그먼트 품질 결정 기법

제안하는 집단 세그먼트 품질 결정 기법은 낭비되는 대역폭을 최소로 하기위한 요청 세그먼트의 수와 품질을 결정한다. 예측 가용 대역폭 $B_{est}[i]$ 는 이전 요청 주기 동안 다운로드 받은 세그먼트들의 정보를 이용하여 식 (1)과 같이 예측한다.

$$B_{est}[i] = \frac{R[i-1] \times \tau \times N[i-1]}{t_{down}} \quad (1)$$

t_{down}은 이전 요청 주기 동안 다운로드 받는데 소요된 시간, $R[i-1]$은 이전에 요청한 콘텐츠의 품질, τ는 세그먼트 길이, 그리고 $N[i-1]$은 이전에 요청한 세그먼트의 수를 의미한다. 전체 세그먼트를 다운로드 받는데 소요된 시간과 다운로드 받은 데이터의 양을 이용하여 가용대역폭을 예측한다. 다음 요청 주기동안 다운로드 할 수 있는 예상 시간 $T_{expect}[i]$는 식 (2)와 같이 계산한다.

$$T_{expect}[i] = T - t_{delay}[\mathrm{i}-1] \quad (2)$$

t_{delay} 는 다운로드 시간과 요청주기의 차이를 의미한다. 이전 요청에 의해 세그먼트를 다운로드 하는 도중 대역폭이 감소하여 다운로드 시간이 요청 주기보다 길어질 경우 다음 요청주기에서 보상하기 위해 다운로드 시간을 조절한다. 예측한 가용 대역폭과 다운로드 할 수 있는 예상 시간을 이용하여 식 (3)과 같이 가용 다운로드 데이터 크기 $S_{avail}[i]$를 예측한다.

$$S_{avail}[i] = B_{est}[i] \times T_{expect}[i] \quad (3)$$

요청 세그먼트의 수 $N[i]$ 의 범위는 버퍼 언더플로우를 방지하기 위해 버퍼 상태 정보와 가용대역폭을 이용하여 식 (4)와 같이 결정한다.

$$N_{th}^{low}[i] \le N[i] \le N_{th}^{high}[i] \quad (4)$$

$N_{th}^{low}[i]$ 는 타겟 버퍼 b_{target} 를 유지하기 위한 최소 필요 세그먼트의 수, $N_{th}^{high}[i]$ 는 오버플로우를 방지하기 위해 최대 버퍼 b_{max}에 도달하기 위해 필요한 세그먼트의 수를 의미한다. $N_{th}^{low}[i]$와 $N_{th}^{high}[i]$는 식 (5), (6)과 같이 계산한다.

$$N_{th}^{low}[i] = \max\left(\frac{b_{target} - b[i]}{\tau}\right) \quad (5)$$

$$N_{th}^{high}[i] = \frac{(b_{max} - b[i])}{\tau} \quad (6)$$

$b[i]$ 는 현재 버퍼 점유량을 의미한다. 최종적으로 요청 세그먼트의 품질과 요청 세그먼트의 수는 요청 세그먼트의

수의 범위와 콘텐츠의 품질 $R[k]$를 이용하여 낭비되는 대역폭이 최소로 되는 값을 결정하고, 식 (7)과 같이 계산한다.

$$(N[i], R[i]) = argmin\,(S_{avail}[i] - R[k] \times \tau \times N[i])$$

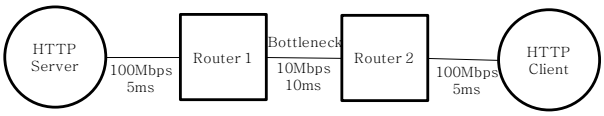

그림 3. 네트워크 환경

3. 실험

제안하는 기법의 성능을 평가하기 위해 NS-3 (Network Simulator) 네트워크 시뮬레이터로 그림 3 과 같이 실험 환경을 구성하였다. HTTP 서버와 HTTP 클라이언트 사이의 병목 구간은 10 Mbps 의 대역폭으로 구성하였다. 콘텐츠의 품질은 각 품질 간의 인코딩 율의 큰 차이를 나타내기 위하여 700, 1400, 2800, 4500, 그리고 9000 Kbps 로 총 5 개의 품질로 구성하였다. 실험은 총 300 초 동안 진행 하였으며, 비교 세그먼트 전송 기술로는 기존 DASH (Dynamic Adaptive Streaming over HTTP)의 전송기술과 제안기법을 비교하였다. 제안하는 세그먼트 품질 조절 기법의 성능을 평가하기 위해서 비교 품질 조절 기법으로는 처리량 기반의 품질 조절 기법 (Conventional)과 버퍼 기반의 품질 조절 기법 BBA (Buffer Based rate Adaptation) 방식을 이용하여 비교를 진행 하였다. 본 실험에서는 요청 주기 T를 8 초로 설정 하였으며, 타켓 버퍼 b_{target}는 20 초, 최대 버퍼의 양 b_{max}는 30 초로 설정하였다.

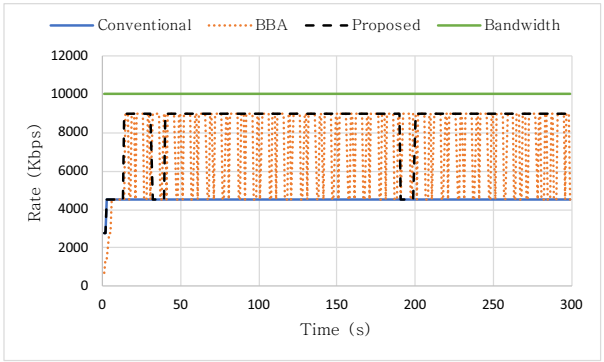

그림 4. 비디오 품질 성능 비교

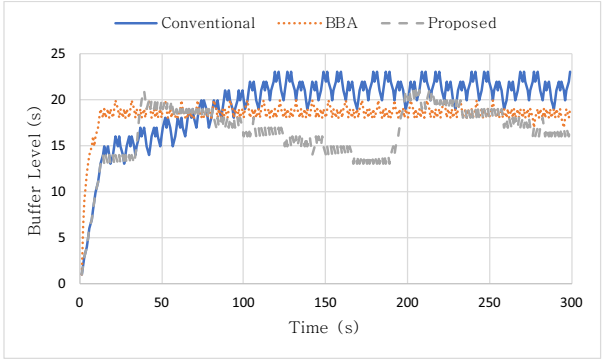

그림 5. 버퍼 점유량 성능 비교

그림 4 와 그림 5 는 제안 기법과 비교 기법의 요청 품질과 버퍼 점유량을 나타낸다. 처리량 기반의 품질 조절 기법은 정상 상태에서 측정한 가용 대역폭 보다 낮은 품질의 콘텐츠를 주기적으로 요청한다. 따라서 4500 Kbps 의 품질을 주기적으로 요청하며, Off 구간에 데이터를 전송하지 않기 때문에 평균 비디오 품질이 낮은 것을 확인할 수 있다. 버퍼 기반의 품질 조절 기법은 버퍼 점유량의 고정된 임계값을 기준으로 품질 변경을 한다. 고정된 임계 값에 의해서 4500, 9000 Kbps 의 품질을 요청하며, 빈번한 품질 변경으로 인해 평균 품질은 향상 되었지만 사용자 체감 품질 (Quality of Experience)이 낮아지는 문제가 발생한다. 반면에 제안 기법은 세그먼트 전송 기법에 의해 Off 구간의 빈도수를 감소시켰으며, 품질 조절 기법으로 인해 가용 대역폭의 낭비를 최소화 하는 품질을 결정하였기 때문에 높은 평균 비디오 품질이 높은 것을 확인할 수 있다. 또한 버퍼 점유량이 낮아지면 빠르게 낮은 품질로 많은 세그먼트를 다운받기 때문에 버퍼가 안정적으로 유지되는 것을 확인 할 수 있다.

4. 결론

본 논문에서는 HTTP 적응적 스트리밍에서 UHD 콘텐츠의 효율적인 대역폭 활용을 위한 세그먼트 전송 기법을 제안하였다. 제안하는 기법은 한번의 요청 메시지에 여러 세그먼트를 동시에 전송하는 집단 세그먼트 전송 방식과 낭비된 대역폭이 가장 적은 품질과 세그먼트의 수를 결정하는 집단 세그먼트 품질 조절 기법으로 구성되어 있다. 제안하는 세그먼트 전송 방식은 Off 구간의 빈도를 줄여주고, 제안하는 품질 조절 기법에 의해 낭비된 대역폭을 최소화 하여 효율적으로 데이터를 전송할 수 있다. 또한 버퍼 상태에 따라 요청 세그먼트의 수를 결정함으로써 버퍼를 안정적으로 관리 할 수 있다. 실험을 통해서 기존 기법에 비해 제안 기법이 높은 평균 비디오 품질을 제공하며 안정적인 버퍼를 점유하는 것을 확인하였다.

감사의 글

이 논문은 2018 년도 정부 (과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00224, UHD 방송콘텐츠 기반 지능형 Dynamic Media 생성, 분배 및 소비 기술 개발)

참고 문헌

[1] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia, “A Survey on Quality of Experience of HTTP Adaptive Streaming,” *IEEE Communications Survey and Tutorials*, vol. 17, no. 1, pp. 469-492, March 2015.

[2] ITU-R Recommendation BT. 2020 “Parameter Values for UHD System for Production and International Programme Exchange,” August 2012.

[3] J. Kua, G. Armitage, and P. Branch, “A Survey of Rate Adaptation Techniques for Dynamic Adaptive Streaming over HTTP,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 1842-1866, March 2017.