



The Development of Realtime Smart Image Mixing System Using GPU Based Video Codec

Seung-Taek Ryoo¹, Young-Kwon Kim², Jae-Khun Chang¹

¹*School of Computer Engineering, HanShin University*

²*Department of Computer Information, Osan University*

A B S T R A C T

As a result of recent advances in GPU programming using graphics hardware, many studies have been focused on the application of parallel and realtime rendering. This paper introduces the development of realtime smart image mixing system using ultra high density video based decoding and encoding techniques. The smart image mixing system implements sequential and streaming based image mixer that consists of video decoding, image resizing and mixing, video encoding step. Sequential based image mixing system decodes whole frames from multiple video inputs and merges extracted synchronous frames from multiple videos into single image. And then, encodes each single image into whole frames and save final result video file. Streaming based image mixing system decodes single frame from multiple video inputs and merges an extracted synchronous frame from multiple videos into single image. And then, encodes a merged image into each frame and repeat the above steps(decoding/mixing/encoding) until last frame from multiple video inputs is extracted. We compare sequential based image mixing method with streaming based image mixing method. In temporal performance, streaming based image mixing method is better than sequential based image mixing method. Suggested method can be used in various fields such as image and game contents generation, UHD image display and network service and video conferencing.

© 2016 KKITS All rights reserved

KEYWORDS : Realistic media, GPU programming, Ultra high definition, Video codec, Smart systems

ARTICLE INFO: Received 7 November 2016, Revised 12 December 2016, Accepted 12 December 2016.

*Corresponding author is with the School of Computer Engineering, HanShin University, 137 Hanshin Univ.-gil

Osan-si, Gyeonggi-do, 18101, KOREA.
E-mail address: jchang@hs.ac.kr

1. 서론

실감 미디어 기술[1]을 현실화하기 위해서는 초고화질 콘텐츠의 제작 기술이 필요하다. 최근 그래픽 하드웨어를 이용한 GPU 프로그래밍[2-5]이 가능하게 되면서 병렬처리 및 실시간 렌더링에 대한 연구가 활발히 이루어지고 있다. 예전까지 불가능하게 여겨졌던 고해상도 영상 처리 기술 및 고급 렌더링 방법들이 GPU를 가진 그래픽 하드웨어의 급속한 발전으로 실시간 렌더링이 가능하게 되었다. 이러한 영상 렌더링 기술을 GPU 기반 비디오 코덱을 이용한 초고화질 영상 제작 기술 개발에 접목한다면 게임, 만화영화, 그래픽 디자인, 디지털 출판, 영상 산업과 같은 다양한 분야에 폭 넓게 활용될 수 있다.

본 연구는 초고해상도 비디오 영상을 실시간으로 처리하는 영상 믹싱 시스템의 개발을 목적으로 한다. 이를 위해 스마트 영상 믹싱 시스템을 순차 처리 방법과 스트리밍 방법으로 나누어 제안하였다. 또한, 각 방법들을 파일 입출력 사용 유무에 따라 렌더링 성능을 비교 평가하였다. 제안된 기술은 향후 영상 및 게임 콘텐츠 제작, 고해상도 영상 디스플레이 및 네트워크 서비스, 화상회의 등에 다양하게 사용될 수 있다.

2. 관련연구

초고화질 영상을 통한 사실감 구현을 목적으로 많은 연구들이 진행되고 있다. 현행 HD 영상보다 해상도가 월등히 증가한 초고화질과 대형화를 구현해 사실감을 높이려는 방식(UHD)을 말한다[6]. 초고화질의 최근 흐름을 보면 SD(Standard Definition), HD(High Definition), UHD(Ultra High Definition) 순으로 발전하고 있다. 해상도에 따라 SD는 720×480(가로×세로), HD는 1920×1080,

UHD는 4K의 경우 3840×2160, 8K의 경우 7680×4320로 구성된다[7-10]. 최근 스마트폰에 탑재되는 디스플레이의 초고화질 경쟁 역시 직접적으로는 픽셀 수의 경쟁이라고 볼 수 있다. 이런 이유로 최근 UHD TV는 초고화질, 대형화 디스플레이에 대한 사용자의 기대를 만족시킬 수 있는 새로운 대안으로 급속히 부상하고 있다.

글로벌 표준화 기구들은 U(H)DTV AV 신호 규격 및 부호화/전송 기술에 대한 표준화 작업을 진행 중이며, 일본의 방송사 및 제조사를 중심으로 U(H)DTV에 대한 연구개발이 주도되고 있다. 일본의 슬림형TV 매출 점유율이 삼성과 LG에 뒤처지면서 일본 가전업계 내에선 TV산업 자체의 붕괴를 우려하는 위기의식 팽배하고 있다. 미국은 방송분야에 있어서 U(H)DTV 기술개발과 적용이 활발하지 않지만, 영화산업 부분은 4K이상의 초고화질 영상 디지털 시네마 표준을 수립하고 4K 영화제작, 디지털 4K 영화관을 보급하고 있다.

국내 UHD관련 연구 및 기술개발은 초기단계이다. 각방송사, 방송통신위원회, 기업 등 다양한 기업에서 UHD TV 기술을 개발 선도하고 있는 추세이다. 2011년부터 한국전자통신연구원(ETRI)을 중심으로 UHD TV와 무안경 다시점 3D TV를 동시에 지원할 수 있는 부호화 및 전송 핵심 기술 개발 과제를 진행 중에 있다. ETRI는 케이블 기반의 4K-UHDTV 전송실험을 실시하였고 이를 토대로 2013년 7월부터 세계 최초 UHD TV 시범방송을 개시하였다. 8K UHDTV는 2018년부터 본 방송이 시작될 예정이며 그전에 현재 4K UHD 관련서비스가 우선적으로 시작되고 있다.

3. 스마트 영상 믹싱 시스템

3.1 전체 시스템 구성

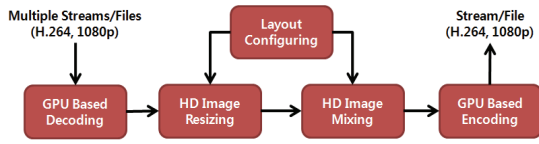


그림 1. 스마트 영상 믹싱 시스템의 구성
Figure 1. The overview of smart image mixing system

초고화질 영상을 사용하여 스마트 영상 믹싱 시스템을 개발하기 위해서는 인코딩/디코딩 단계를 실시간적으로 처리해야 한다. 이를 위해 본 연구에서는 GPU 기반 디코딩 기능, 영상 리사이징 기능, 영상 믹싱 기능, 레이아웃 설정 기능, GPU 기반 인코딩 기능을 개발하였다.

GPU 기반 디코딩 단계에서는 인코딩(H.264)된 상태의 스트림이나 파일을 영상처리가 가능한 데이터로 복호화한다. 코덱에 대한 복호화 기능은 하드웨어 가속 처리가 가능하다. 영상 리사이징 단계에서는 영상의 크기를 원본 영상 이하의 크기로 축소하고 변경된 영상은 믹싱 단계의 입력 영상으로 사용된다. 영상 믹싱 단계에서는 다수의 영상을 설정된 레이아웃으로 하나의 영상을 생성한다. 레이아웃 설정(Layout Configuring) 단계에서는 비디오 영상을 믹싱하기 위한 다중 비디오 영상의 레이아웃을 설정한다. 레이아웃 정보는 믹싱과 리사이징에서 활용된다. GPU 기반 인코딩 단계에서는 스트림이나 파일을 생성하기 위해 데이터를 부호화한다. 인코딩용 코덱은 H.264이고 하드웨어 가속 처리가 가능하다.

3.2 영상 믹싱 시스템의 처리 방식

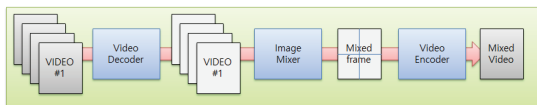


그림 2. 스마트 영상 믹싱 시스템의 처리과정
Figure 2. The process of smart image mixing system

본 연구에서는 스마트 영상 믹싱 시스템을 위해 비디오 디코딩, 영상 리사이징 및 믹싱, 비디오 인코딩을 개발하였고 순차처리 기반 영상 믹서(모든 프레임)와 스트리밍 기반 영상 믹서(한 프레임)로 나누어 개발하였다.

순차처리 기반 영상 믹싱 시스템은 비디오 영상으로부터 디코딩을 하여 모든 프레임들을 추출하고 다수의 비디오 영상으로부터 추출된 모든 프레임들을 하나의 영상으로 믹싱한다. 믹싱된 모든 영상 프레임들을 인코딩을 하여 비디오 파일로 저장한다. 스트리밍 기반 영상 믹싱 시스템은 비디오 영상으로부터 디코딩을 하여 하나의 프레임들을 추출한다. 다수의 비디오 영상으로부터 추출된 각각의 프레임들은 하나의 영상으로 믹싱한다. 믹싱된 영상 프레임들을 한프레임씩 인코딩을 하여 비디오 파일로 저장한다. 디코딩, 믹싱, 인코딩 과정을 마지막 비디오 영상이 추출될 때까지 반복한다.

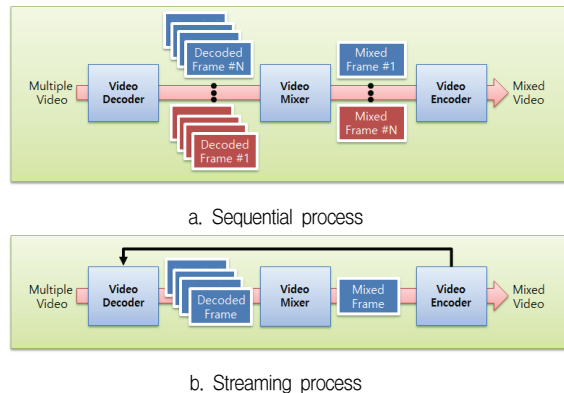


그림 3. 순차 처리 기반 영상 믹싱과 스트리밍 기반 영상 믹싱 시스템 비교

Figure 3. The comparison of sequential process and streaming process

3.3 비디오 디코딩

디코딩 단계는 CPU 기반 디코딩과 GPU 기반

디코딩으로 나뉜다. CPU 기반 디코딩에서는 OpenCV[11]를 이용하여 인코딩된 상태의 파일을 영상처리가 가능한 데이터로 모든 프레임들을 복호화하고 GPU 기반 디코딩에서는 CUDA[12] 기반 NVCUVID를 이용하여 인코딩된 상태의 파일을 영상처리가 가능한 데이터로 모든 프레임들을 복호화한다.

NVCUVID[13]을 이용한 디코딩 처리 과정을 자세히 살펴보면 아래와 같다. 먼저 CUDA 비디오 디코더 API를 이용하여 입력 비디오를 파싱한다. 파싱된 비디오는 NVCUVID API를 이용하여 YUV 포맷의 프레임으로 비디오 디코딩한다. YUV 포맷의 프레임을 RGBA 영상 포맷으로 변환한다. 칼라 모델 변환된 프레임은 OpenGL[14] 서페이스를 이용하여 RGBA 서페이스에 매핑한다. 마지막으로 RGBA 서페이스를 텍스처 메모리에 저장후 CUDA 서페이스 렌더링을 사용하여 비디오 디스플레이를 하거나 OpenCV를 이용하여 비트맵으로 파일을 저장한다.

3.4 영상 믹싱

영상 믹싱 단계에서는 다수의 영상들의 복호화된 프레임들의 크기를 조절하여 하나의 영상으로 통합한다. 이 단계도 CPU 기반 영상 믹싱과 GPU 기반 영상 믹싱으로 나뉜다. CPU 기반 영상 믹싱에서는 OpenCV를 이용하여 각각 영상의 크기를 조절하고 통합한다.

GPU 기반 영상 믹싱에서는 RGBA 서페이스를 OpenGL을 이용하여 파일 저장없이 텍스처 자체로 바로 사용하거나 비트맵 파일로 저장후 연동할 수 있다. 또한, RGBA 서페이스를 OpenCV를 이용하여 파일 저장없이 메모리 블록을 IplImage로 생성하여 연동하거나 비트맵 파일로 저장후 연동할 수 있다.

3.5 비디오 인코딩

인코딩 단계는 디코딩 단계와 같이 CPU 기반 인코딩과 GPU 기반 인코딩으로 나뉜다. CPU 기반 인코딩에서는 OpenCV를 이용하여 파일을 생성하기 위해 믹싱된 모든 프레임들을 부호화한다. GPU 기반 인코딩에서는 GPU 기반 NVENC 이용하여 파일을 생성하기 위해 믹싱된 모든 프레임들을 부호화한다.

NVENC[15]를 이용한 인코딩 처리 과정은 코덱, 크기 및 비트레이트 등의 비디오 프로필을 세팅하고 디코딩 단계와 반대로 RGBA 이미지를 YUV 포맷의 프레임으로 변환한다. 이렇게 변환된 프레임은 H.264 코덱을 이용하여 비디오 파일로 저장한다.

4. 구현결과

본 연구에서는 스마트 영상 믹싱 시스템의 성능을 평가하기 위해 제안된 GPU 기반 방법들을 비교하였다. CPU 기반 인코딩/디코딩 방법은 GPU 기반 방법에 비해 실시간 렌더링이 어려워 성능 평가에서 제외하였다.

표 1. 순차처리 기반 영상 믹싱 시스템의 성능 (fps)
Table 1. The performance of sequential based Image Mixing system (fps)

| GPU 기반 순차 처리 영상 믹서 | | 720p | 1080p |
|---------------------------|-----|-------|-------|
| OpenCV (with File I/O) | 디코딩 | 6.12 | 2.62 |
| | 믹싱 | 21.85 | 9.58 |
| | 인코딩 | 8.52 | 4.26 |
| | 전체 | 3.06 | 1.39 |
| OpenGL (with File I/O) | 디코딩 | 63.78 | 22.79 |
| | 믹싱 | N/A | N/A |
| | 인코딩 | 25.66 | 18.32 |
| | 전체 | 18.30 | 10.16 |

스마트 영상 믹싱 시스템의 성능을 평가하기 위한

실험 환경은 윈도우8(64bit), CPU i7-2600(3.4GHz), RAM 8.0GB, GPU geforce GTX760 사용하였다. 실험 영상은 720p (1280x720, 24 fps, 약 11초), 1080p (1920x1080, 24.98 fps, 약 9초)를 사용하였다.

표 2. 스트리밍 기반 영상 믹싱 시스템의 성능 (fps)
Table 2. The performance of streaming based Image Mixing system (fps)

| GPU 기반 스트리밍 영상 믹서 | | 720p | 1080p |
|-----------------------------|-----|--------|-------|
| OpenCV버전 (with File I/O) | 디코딩 | 6.98 | 3.05 |
| | 믹싱 | 65.66 | 32.35 |
| | 인코딩 | 107.52 | 62.79 |
| | 전체 | 5.96 | 2.67 |
| OpenCV버전 (w/o File I/O) | 디코딩 | 52.36 | 23.10 |
| | 믹싱 | 106.58 | 45.79 |
| | 인코딩 | 133.93 | 66.04 |
| | 전체 | 27.82 | 12.46 |
| OpenGL버전 (with File I/O) | 디코딩 | 72.21 | 29.51 |
| | 믹싱 | N/A | N/A |
| | 인코딩 | 133.42 | 60.12 |
| | 전체 | 45.96 | 19.62 |
| OpenGL버전 (w/o File I/O) | 디코딩 | 89.23 | 42.49 |
| | 믹싱 | N/A | N/A |
| | 인코딩 | 134.36 | 64.44 |
| | 전체 | 52.47 | 25.33 |

본 연구에서는 순차처리 기반 영상 믹싱 시스템의 성능을 측정하기 위해 <표 1>과 같이 OpenCV 기반 메모리 사용과 OpenGL 기반 메모리 사용으

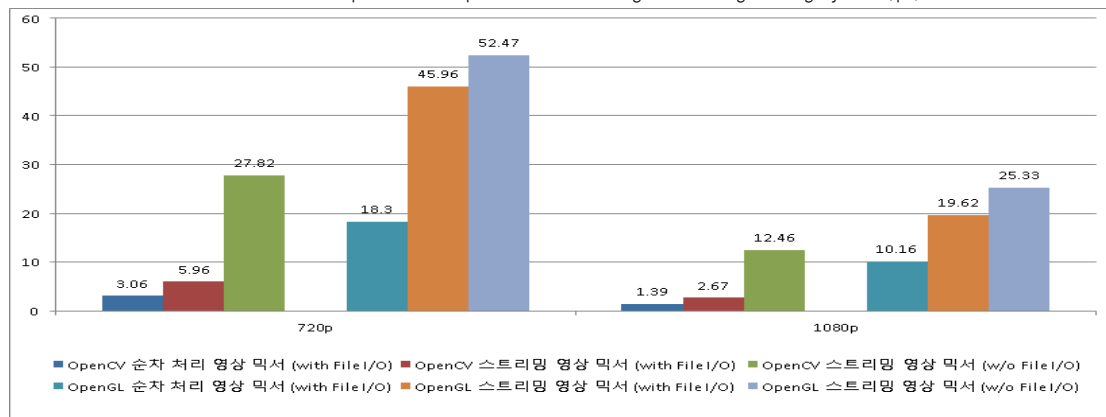
로 나누어 측정하였다.

스트리밍 기반 영상 믹싱 시스템의 성능을 측정하기 위해 <표 2>와 같이 OpenCV 기반 메모리 사용, OpenCV 기반 메모리와 파일 저장 사용, OpenGL 기반 메모리 사용, OpenGL 기반 메모리와 파일 저장 사용으로 나누어 측정하였다.

실험 결과를 보면 디코더가 성능에 대한 주요 병목지점을 알 수 있으며 <표 3>과 같이 파일 입출력없이 OpenGL을 사용할 경우 가장 좋은 성능을 얻을 수 있었다. 이는 OpenGL을 사용함으로써 디바이스와 호스트 사이의 데이터 교환 횟수를 줄였기 때문이다. 또한, 순차처리 버전에 비해 스트리밍 버전의 성능이 더 뛰어난 것을 알 수 있다. 이론적으로는 각 모듈의 문맥 교환(context switching)이 빈번하게 발생하는 스트리밍 버전이 더 느려야 하지만, 스트리밍 버전은 파일 입출력(File I/O)을 필요로 하지 않기 때문에 순차처리 버전에 비해 OpenCV 사용시 약 8~9배, OpenGL 사용시 약 2배 빠른 결과를 얻었다.

전체 수행 시간은 디코더 및 인코더의 사전 CUDA 세팅 시간을 포함하고 있기에, 세 모듈(디코더/믹서/인코더)의 총 fps보다 느리게 수행된다. 실험에 사용된 영상들은 비교적 길이가 짧기 때문에

표 3. 순차처리와 스트리밍 기반 영상 믹싱 시스템의 비교 (fps)
Table 3. The comparison of sequential and streaming based Image Mixing system (fps)



사전 처리과정의 소요 시간이 전체 성능에 크게 영향을 주었지만, 일반적인 비디오 영상은 길이가 훨씬 길기 때문에, 전체 시간은 세 모듈의 합에 수렴해갈 것임을 예측할 수 있다.

인코더와 믹서의 성능이 순차처리버전과 스트리밍 버전에서 큰 차이가 있는데, 이것은 순차처리 버전에서만 인코더와 믹서에 파일 입출력이 내장되어 있기 때문이다. 순차처리 버전은 파일 입출력이 필수적으로 요구되어 모든 모듈에 파일 입출력이 내장되어있다. 반면 스트리밍 버전은 중간 과정을 보기 위한 용도로 파일 입출력 옵션을 사용하였기 때문에, 디코더에만 파일 입출력을 내장한다.

실험에 사용된 GPU는 최신 사양이 아님에도 불구하고, CUDA와 OpenGL 환경에서 실시간에 준하는 성능을 얻었다. 따라서 최신 GPU 사용시, 비약적인 성능 향상이 예상된다.

5. 결론

본 연구는 GPU 기반 비디오 코덱을 이용한 스마트 영상 믹싱 시스템 개발을 목적으로 한다. 이를 위해 순차처리 기반 영상 믹서(모든 프레임)와 스트리밍 기반 영상 믹서(한 프레임)으로 나누어 개발하였다. 또한, 영상 믹싱시 OpenCV 기반 메모리와 OpenGL 기반 메모리를 이용하였다. 스트리밍 기반 영상 믹서는 파일 저장없이 메모리 블록을 직접 사용하는 경우와 파일 저장 후 연동하는 경우로 나누어 구현하였다. 본 연구에서 제안된 OpenGL 메모리 사용한 GPU 기반 스트리밍 영상 믹싱 방법이 순차처리 방법에 비해 약 17~18배의 성능 향상이 되었음을 확인할 수 있었다.

제안된 스마트 영상 믹싱 기술은 향후 영상 및 게임 콘텐츠 제작, 고해상도 영상 디스플레이 및 네트워크 서비스, 화상회의등에 다양하게 사용될 수 있을 것이다.

References

- [1] W. C. Seo, G. I. Kim, and J. S. Jung, *The trend of Real media techniques in smart era*, The Korean Institute of Communications and Information Science, Vol. 30, No. 5, pp. 78-87, 2013.
- [2] H. Nguyen, *GPU gems 3*, Addison-Wesley Professional, 2007.
- [3] G. Singer, *The history of the modern graphics processor*, Techspot, Mar. 2013.
- [4] J. D. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Krüger, A. E. Lefohn, and T. J. Purcell, *A survey of general-purpose computation on graphics hardware*, Computer Graphics Forum, Vol. 26, Issue 1, pp. 80-113, 2007.
- [5] Graphics Processing Unit (GPU), Nvidia Corporation, Sep. 2016.
- [6] S. R. Kim, *The trend of ultra UHD image industry*, Journal of Communication & Radion Spectrum, Special Issue, pp. 54-61, 2012.
- [7] A. Thomas, *Just how useful is 2160p aka 4K?*, Sep. 2016.
- [8] A. Cotton, *Defining the future of television*, BBC, Nov. 2015.
- [9] Leading Television Industry Players Line Up To Support '4K Ultra HD', Consumer Electronics Association, 2014.
- [10] Y. Shishikui, K. Iguchi, S. Sakaida, K. Kazui, and A. Nakagawa, *Development of high performance video codec for super hi-vision*, 65th NAB Broadcast Engineering Conference, pp. 234-239, 2011.
- [11] P. Joshi, D. M. Escrivá, and V. Godoy, *OpenCV By example*, PACKT Publishing, Jan. 2016.

[12] J. Sanders, and E. Kandrot, *CUDA by example: An introduction to general-purpose GPU programming*, Addison-Wesley Professional, 2010.

[13] *Video Decoder Programming Guide*, Nvidia Corporation, Jun. 2016.

[14] G. Seller, R. S. Wright Jr, and N. Haemel, *OpenGL superbible: Comprehensive tutorial and reference*, Addison-Wesley Professional, 2013.

[15] *Video Decoder Programming Guide*, Nvidia Corporation, Jun. 2016.

기반 영상 믹싱 방법과 스트리밍 기반 영상 믹싱 방법을 비교하여 보면 스트리밍 기반 영상 믹싱 방법이 순차 기반 영상 믹싱 방법보다 더 성능이 우수함을 알 수 있다. 이러한 기술은 향후 영상 및 게임 콘텐츠 제작, 고해상도 영상 디스플레이 및 네트워크 서비스, 화상회의등에 다양하게 사용될 수 있다.

GPU 기반 비디오 코덱을 이용한 실시간 스마트 영상 믹싱 시스템의 개발

류승택¹, 김영권², 장재건¹

¹한신대학교 컴퓨터공학부

²오산대학교 컴퓨터정보학과

요 약

최근 그래픽스 하드웨어를 이용한 GPU 프로그래밍이 가능하게 되면서 병렬처리 및 실시간 렌더링에 대한 연구가 활발히 이루어지고 있다. 본 연구는 초고해상도 비디오 영상 기반 디코딩/인코딩 기술을 이용한 스마트 영상 믹싱 시스템 개발을 목적으로 한다. 스마트 영상 믹싱 시스템은 비디오 디코딩, 영상 리사이징 및 믹싱, 비디오 인코딩 단계로 구성된 순차처리와 스트리밍 기반 방식으로 구현하였다. 순차처리 기반 영상 믹싱 시스템은 다수의 비디오 영상으로부터 디코딩을 하여 모든 프레임들을 추출하고 다수의 비디오 영상으로부터 추출된 프레임들을 하나의 영상으로 통합한다. 믹싱된 모든 영상 프레임들을 인코딩을 하여 비디오 파일로 저장한다. 스트리밍 기반 영상 믹싱 시스템은 비디오 영상으로부터 디코딩을 하여 하나의 프레임들을 추출한다. 다수의 비디오 영상으로부터 추출된 각각의 프레임들은 하나의 영상으로 믹싱한다. 믹싱된 영상 프레임들을 한프레임씩 인코딩을 하여 비디오 파일로 저장한다. 디코딩, 믹싱, 인코딩 과정을 마지막 비디오 영상이 추출될 때까지 반복한다. 순차

감사의 글

본 논문은 중소기업청에서 지원하는 2015년도 산학연협력 기술개발사업(기업부설연구소 신규설치)(No. C0261447)의 연구수행으로 인한 결과물임을 밝힙니다.



Seung Taek Ryoo received the B.S. and M.S. degrees in Computer Science Engineering from ChungAng University in 1996 and 1998, respectively. He had received his Ph.D.

in Department of Image Engineering, Graduate School of Advanced Imaging Science, Multimedia and Film from ChungAng University in 2002. He is currently a professor in School of Computer Engineering, Hanshin University, Korea. His interests are computer graphics, image-based rendering, realtime rendering, non-photorealistic rendering.

E-mail address: stryoo@hs.ac.kr



Young Kwon Kim received his Ph.D. in Department of Computer Science & Engineering from ChungAng University in 2009. He is currently a professor in

Department of Computer Information, Osan University, Korea. His interests are artificial intelligence, embedded system, ubiquitous.

E-mail address: ykkim@osan.ac.kr



Jae Khun Chang received his B.S. degree from Hanyang University in 1985. He had received his M.S. degree from Department of Computer and Information

Science in New Jersey Institute of Technology, and Ph.D. from Department of Computer Science in University of South Carolina in 1989 and 1997 respectively. He is currently in School of Computer Engineering, Hanshin University, Korea. His interests are computer vision, ITS, image processing, pattern recognition, and motion tracking.

E-mail address: jchang@hs.ac.kr