

# 깊이 맵 기반의 실시간 가상 영상합성 시스템의 설계 및 구현

이혜미\* · 류남훈\*\* · 노관승\*\*\* · 김응곤\*\*\*\*

The Design and Implementation of Real-time Virtual Image Synthesis System of Map-based Depth

Hye-Mi Lee\* · Nam-Hoon Ryu\*\* · Gwhan-Sung Roh\*\*\* · Eung-Kon Kim\*\*\*\*

## 요 약

영상물을 완성하기 위해서는 실사 배우의 동작을 캡처하여 가상 환경과 합성하는 과정을 거치게 된다. 이때 제작비용의 과다소요 또는 후처리 기술의 부재로 인하여 수작업으로 제작하는 것이 일반적이다. 배우는 크로마키 가상스튜디오에서 상상력에 의존하여 연기를 하게 되는데, 실존하지 않는 물체와의 충돌 및 반응을 고려하여 움직여야 한다. CG를 통한 합성 과정에서 배우의 움직임과 가상 환경이 어긋나는 경우는 원본 영상을 폐기하고, 재촬영해야 하는 문제가 발생한다. 본 논문에서는 재촬영의 비율을 줄이고, 제작기간 단축 및 제작비용절감을 위한 깊이 기반의 실시간 3D 가상 영상합성 시스템을 제안하고 구현한다. 가상 배경과 3D 모델, 실사 배우를 실시간으로 합성하여 상호간의 충돌이나 반응을 촬영 현장에서 판단할 수 있으므로 배우의 잘못된 위치나 연기를 개선할 수 있게 한다.

## ABSTRACT

To complete an image, it is needed to go through the process to capture the actual actor's motion and compose it with virtual environment. Due to the excessive cost for production or lack of post-processing technology, however, it is mostly conducted by manual labor. The actor plays his role depending on his own imagination at the virtual chromakey studio, and at that time, he has to move considering the possible collision with or reaction to an object that does not exist. And in the process of composition applying CG, when the actor's motion does not go with the virtual environment, the original image may have to be discarded and it is necessary to remake the film. The current study suggested and realized depth-based real-time 3D virtual image composition system to reduce the ratio of remaking the film, shorten the production time, and lower the production cost. As it is possible to figure out the mutual collision or reaction by composing the virtual background, 3D model, and the actual actor in real time at the site of filming, the actor's wrong position or acting can be corrected right there instantly.

## 키워드

Image Composition, Chromakey, Kinect, Depthmap  
영상합성, 크로마키, 키넥트, 깊이 맵

\* 순천대학교 컴퓨터학과(lhrooh@sunchon.ac.kr)

\*\* (주)위더스비즈 대표이사(ceo@withusbiz.com)

\*\*\* 온유기술연구소 대표(onewtech1240@gmail.com)

\*\*\*\* 교신저자(corresponding author) : 순천대학교 컴퓨터공학과 교수(kek@sunchon.ac.kr)

접수일자 : 2014. 09. 26

심사(수정)일자 : 2014. 10. 20

게재확정일자 : 2014. 11. 10

## I. 서론

영화를 비롯한 방송 매체에서 영상을 제작하기 위해 디지털 기술을 사용하는 것은 보편화된 현상이다. 창의적인 아이디어를 실현시켜 줄 수 있는 컴퓨터 그래픽 기술은 실사와 가상 환경을 합성시켜 제작 공간에 대한 제약을 없애고, 다양한 콘텐츠 생성을 가능하게 하였다. 실사 촬영이 어려운 영상물의 경우는 컴퓨터 그래픽스 기술을 이용하여 실제 연기와 합성하게 되는데, 가상의 배경이나 객체는 연기와 이음새 없이 결합이 가능해야 하고, 영상의 이질감은 최소화되어야 한다. 거대 자본이 투입된 현실감 있는 고품질 영상은 시청자의 요구 수준을 높이게 되었고, 이를 극복하기 위해 저비용 고효율의 영상 제작 기술의 개발이 절실히 필요한 시점이다.

본 논문에서는 실시간으로 2D 배경 및 3D 가상 모델이 적용된 통합 가상환경과 배우를 합성하여 촬영과 동시에 예상 결과화면을 모니터링 할 수 있는 시스템을 설계하고 구현하고자 한다. 배우와 3D 가상 객체의 충돌 및 반응을 촬영 중에 판단할 수 있으므로 배우의 잘못된 동작이나 위치를 수정하면서 촬영을 끝마칠 수 있게 된다. II장에서는 영상 합성에 필요한 제반기술에 대해 설명하고, III장에서는 깊이 정보 기반의 실시간 가상 영상합성 시스템에 대해 설명하며, IV장에서는 구현 결과를 살펴보고, V장에서는 결론 및 기대효과를 제시한다.

## II. 관련연구

### 2.1. 가상 스튜디오 합성기법

영상물을 완성하기 위해서는 실제 배우의 동작을 캡처하여 가상 환경과 합성하는 과정을 거치게 되는데 이때 비용의 문제 및 후처리 기술의 부족으로 인하여 합성 과정을 수작업으로 제작하는 것이 일반적이다. 프리프로덕션(Preproduction) 과정을 통해 실사의 배우와 가상의 객체가 어긋나지 않도록 사전 준비를 하지만, 실사 촬영 시에는 감독이 결과 영상을 확인할 수 없다. 그렇기 때문에 포스트 프로덕션(Post Production) 과정에서 추출된 배우 객체의 움직임과 합성될 3D 가상 환경이 제대로 일치하지 않는 경우가 자

주 발생한다.

포스트 프로덕션 과정에서는 크로마키(Chromakey)나 로토스코핑(Rotoscoping)에 의해 추출된 실제 배우 객체와 3D 가상 환경을 1프레임 단위로 합성하는 작업을 거치게 된다. 크로마키 환경에서 배우가 취하는 액션은 순수 상상력에 의존하여 결정되고, 실존하지 않는 물체와의 충돌 및 상호작용을 고려하여 움직인다. 그러나 배우의 위치나 시선, 감정 등이 미묘하게 어긋나는 경우는 크로마 스튜디오(Chroma Studio)[1]에서 다시 촬영을 해야 하고, 재촬영된 영상물에서 프레임 당 합성하는 작업을 반복해야 한다. 이때 재촬영을 위해서는 앞서 촬영한 환경(카메라의 간격 및 각도, 조명 환경, 구조물 등)과 동일해야 하며, 배우와 엑스트라, 스태프 등 모든 인력이 재동원되어야 한다는 문제가 발생한다. 게다가 수작업을 하는 동안 시간이 지나버리게 되어 이전의 감정을 그대로 되살리기에 는 무리가 있어 연기의 완성도가 떨어질 확률 또한 높아지게 된다.

### 2.2. 전경 객체 분리 기법

다양한 시각 효과를 위해 원본 영상에서 특정 영역을 추출하는 시각 효과 합성 방법으로는 크로마키 기법과 로토스코핑 기법이 있다. 크로마키는 컴퓨터가 배경에서 관심영역을 쉽게 분리할 수 있도록 단일 색상으로 구성된 스튜디오에서 피사체를 촬영한 후 새로운 배경플레이트와 관심영역을 중첩하는 기술이다. 단일 색상으로는 주로 청색을 사용하기 때문에 블루스크린 합성 기법이라고도 하지만 카메라를 통해 입력된 영상에서 특정 색상 영역을 임의로 지정할 수 있으므로 색상의 제한은 없다. 그림 1은 크로마키에 의해 가상 환경이 적용된 모습을 보여준다.



그림 1. 크로마키에 의한 가상 영상합성  
Fig. 1 Virtual image synthesis by chromakey

로토스코핑 기법은 단일 색상을 배경으로 하지 않은 상태에서 피사체를 고립시켜야 할 때 필요한 관심

영역을 영상에서 수작업으로 분리해내는 기법을 말한다. 각 프레임 별로 직접 수작업을 통해 분리 및 합성이 이루어지므로 매우 느리고, 노동집약적이지만 상황에 따라 유일한 해결책이 되기도 한다[2].

이러한 전경 객체 분리 기법은 전경과 배경의 2-layer로만 분리된다. 크로마키 배경은 영역 전체가 가상 영상으로 대체되는 구조이므로 부분적인 대체는 불가능하다. 3-layer 더블 크로마키는 전경인 피사체와 배경인 가상환경에 3D 모델을 추가로 합성할 수는 있지만 실시간 z-buffering이 적용되지 않기 때문에 배경-피사체-3D 모델의 앞뒤 정렬 순서는 뒤바뀌지 않는다. 멀티플레인 합성(Multiplane Compositing)으로는 2D와 3D 작업을 연결할 수는 있지만 2D 이미지들이 3D 공간에 Z축으로 나열되어 있으면서 가상 카메라의 이동을 통해 장면이 보이게 된다[2]. 이와 같은 기술은 추후 포스트 프로덕션 과정에서 수작업으로 이루어지는 과정이므로 3D 객체와 함께 연기를 해야 하는 영상물을 촬영할 때 합성 결과물을 실시간으로 시각화하기에는 적합하지 않다.

### III. 실시간 가상 영상합성 시스템

이 장에서는 크로마키의 단편적인 영상 합성을 배제하고, 깊이 정보를 융합하여 2D 배경을 대체하고, 실사 객체와 3D 객체가 상호 간에 실시간으로 대응할 수 있는 영상합성 시스템을 설계하고 구현한다. 이 시스템은 스토리나 시나리오에 따른 영상물을 제작하는데 그 목적이 있으므로 미리 가상 환경 및 3D 객체의 애니메이션과 같은 제반 작업이 선행되어 있어야 한다. 2-layer 합성 모듈, 3D 가상공간 생성 모듈, 2D-3D 가상 영상합성 엔진으로 구성되어 있으며, 그림 2는 본 시스템의 전체 구성도를 나타낸다.

가상 스튜디오에서 움직이는 배우의 형태를 추출하되 깊이 카메라를 이용하여 공간에 대한 깊이 값과 배우 영역에 대한 깊이 값을 획득한다. 동시에 RGB 카메라를 통해 배우 영역에 대한 컬러 영상을 검출하고, 합성될 배우 영역만을 분리해낸다. 깊이 값이 포함된 컬러 영상으로 분리된 배우 영역은 가상공간에서 2D 배경이미지, 3D 모델과 함께 가상환경에서 합성된다.

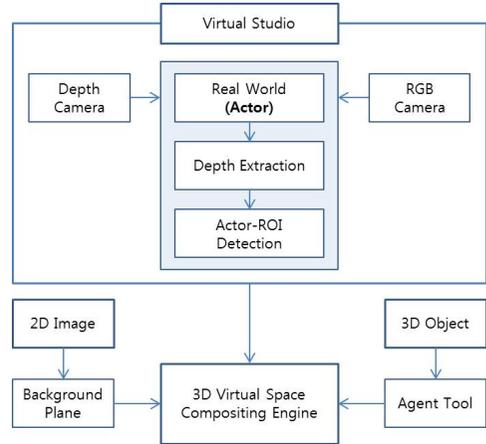


그림 2. 전체 시스템 구성도  
Fig. 2 Whole system structure

#### 2.1. 2-layer 합성 모듈

기존의 크로마키잉(Chroma-Keying)에 의한 전경과 배경의 영역 분리 및 배경 중첩(Superimpose)의 경우는 깊이 버퍼가 생성되지 않는다. 3D 모델을 연동하기 위해서는 3차원 공간 데이터가 생성되어야 하므로 이를 위해서 깊이 키잉(Depth-Keying) 방식을 통해 전경과 배경을 분리한다.

정밀한 깊이 데이터를 획득하기 위한 하드웨어는 고가의 장비이므로, 저렴한 비용으로 깊이 영상과 컬러 화상을 동시에 획득 하고자 하는 경우 키넥트를 이용하여 RGB 영상과 깊이 스트림(Depth Stream)을 입력받는다. 키넥트란 3차원 공간 데이터를 생성하고, 적외선 패턴 인식을 통해 입력된 영상의 깊이를 측정하여 다양한 사용자 경험을 제공하는 디바이스이다 [3-5]. 그림 3은 2-layer 합성 모듈의 파이프라인을 나타낸다.

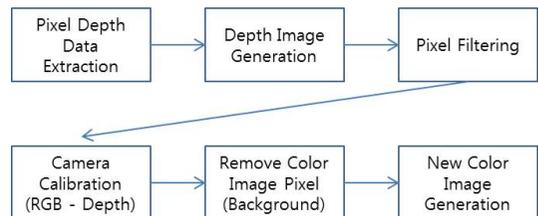


그림 3. 2-Layer 합성 모듈의 파이프라인  
Fig. 3 Pipeline of the 2-layer composition module

먼저 전경을 별도의 객체로 추출하기 위해서는 입력된 원시(Raw) 깊이 영상에서 각각의 픽셀 깊이 값을 알아야 한다. 각 픽셀의 깊이 값은 영상의 시야범위의 물리적인 거리와 대응하여 매핑된다. 각 픽셀의 깊이 정보를 저장하는 비트 필드는 총 16비트이고, 그 중 깊이 비트(Depth Bits)는 13비트를 사용하며, 플레이어 인덱스(Player Index)는 3비트를 사용한다. 깊이 비트(Depth Bits) 13비트 중 12비트는 각 픽셀의 깊이 정보에 사용되고, 1비트는 깊이 측정의 불가여부에 사용된다. 플레이어 인덱스(Player Index)는 인간의 형태를 감지하기 위한 정보로 전경을 추출하기 위한 값에 해당한다. 그림 4는 깊이 비트의 레이아웃을 나타낸다[6].

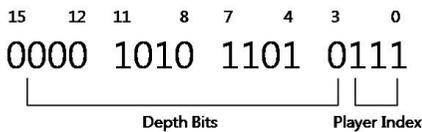


그림 4. 깊이 비트의 레이아웃  
Fig. 4 Layout of the depth bits

생성된 깊이 이미지에서 전경을 분리하기 위해서는 필터링 작업이 수행되어야 한다. 키넥트는 0 ~ 4095까지의 깊이 값을 리턴하며, 각 수치는 길이 단위 mm에 대응한다. 2-layer 합성 모듈에서는 전경의 위치에 대한 임계값을 800 ~ 4000으로 설정하고, 깊이 임계값이 4000을 초과하는 픽셀은 필터링하여 후면으로 처리한다. 그림 5의 (a)는 RGB 카메라를 통해 입력된 컬러 영상을 나타내고, (b)는 공간의 깊이를 시각화시켜주는 깊이 영상을 나타낸다.

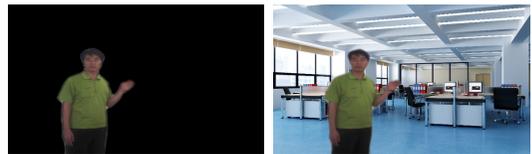


(a) RGB 이미지 (a) RGB Image  
(b) 깊이 이미지 (b) Depth Image

그림 5. 깊이 이미지에서 전경을 분리하는 과정  
Fig. 5 The process of separating the foreground from the depth image

깊이 카메라를 통해 입력된 깊이 이미지의 픽셀은

RGB 카메라를 통해 입력된 컬러 스트림과 1:1로 변환되지 않는다는 문제가 있는데, 그 이유는 두 카메라의 위치가 다르기 때문이며, 카메라 캘리브레이션 [7-8] 작업을 통해 깊이 이미지의 픽셀 위치를 컬러 이미지의 픽셀 위치로 변환한다. 이 후 전경이 아닌 부분의 픽셀은 제거하고, 새로운 컬러 이미지를 생성하여 후면에 해당하는 부분에 컬러 이미지를 드로잉한다. 이 때 전경 주위 픽셀들의 스무딩(Smoothing) 처리를 하지 않으면 경계 주변으로 잡음이 생기게 되므로, Edge를 보정하여 잡음을 제거한다[9]. 그림 6의 (a)는 후면을 필터링한 후의 전경 RGB 영상이며, (b)는 디바이스를 통해 입력된 실사 영상의 후면을 2D 배경 이미지로 변경하고, 전경을 합성한 영상이다.



(a) 전경 추출 (a) Foreground detection  
(b) 합성 이미지 (b) Synthesis image

그림 6. 깊이 정보 기반의 전경과 배경 영상의 합성  
Fig. 6 Synthesis of the foreground and background image based on the depth information

### 2.2. 3D 가상공간 생성 모듈

배우가 시나리오에 따른 연기를 하기 위해 필요한 3D 가상환경을 구축하는 모듈이다. 3D 객체는 최종 결과물로 합성되기 이전에 사용자가 임의로 확대/축소 및 회전, 이동 등의 제어를 할 수 있도록 에이전트 툴로 로딩된다. 에이전트 툴에서는 3D 객체에 대한 사용자의 입력 데이터를 저장하고, 가상영상 합성엔진에 전달한다. 그림 7은 트랜스포머 범블비의 3D 가상 모델을 에이전트 툴로 로딩한 후 합성될 위치를 선정하기 위해 좌표값을 임의로 조정하고 있는 모습이다.

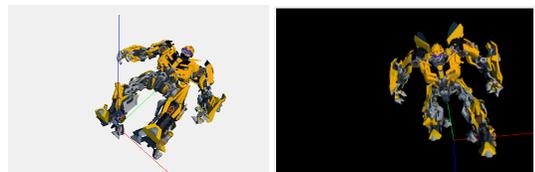


그림 7. 에이전트 툴  
Fig. 7 Agent tool for 3D object

### 2.3. 2D-3D 가상 영상합성 엔진

최종적인 영상의 결과물을 확인하기 위해 필요한 엔진에서는 2D 배경과 3D 가상환경을 하나로 합성하는 과정을 거친다. 2-layer 합성 모듈에서 처리가 완료된 배경 대체 작업과 3D 가상 모델의 렌더링을 위한 공간을 합친다. 원근 투영(Perspective Projection)을 적용하여 원근감을 부여하고, Z-버퍼링을 통해 전경과 3D 가상 모델이 겹치는 부분은 은면(Hidden Surface) 처리하여 앞뒤 정렬 순서를 시각적으로 확인할 수 있도록 하였다. 후면으로 지정된 2D 배경은 Z축을 기준으로 가장 뒤쪽의 화면 전체에 드로잉 된다. 추출된 전경 배우의 모습과 3D 가상 모델의 깊이 테스트는 실시간으로 계산되어 두 객체의 움직임에 따라 Z-버퍼링이 결정된다.

## IV. 실험 및 결과

본 시스템은 가상 영상물의 촬영 시 재촬영의 비율을 줄이기 위한 기술이므로 영상 속에 등장하는 가상 객체들에 맞춰 실사 배우의 위치나 시선이 어긋나지 않도록 하는데 그 목적이 있다. 배우는 최종 결과물에 배치되는 3D 객체가 존재하지 않는 가상 스튜디오에서 연기를 하게 되지만, 영상 속에서는 가상 환경과 3D 객체가 함께 등장한다. 따라서 가상 환경에 맞춘 실사 배우의 움직임이나 시선의 흐름을 관찰할 수 있다. 그림 8은 가상 배경과 실사 배우, 3D 가상 모델이 실시간으로 움직이는 상황을 나타낸다.



그림 8. 최종 실험 결과화면

Fig. 8 Screen of the final implementation results

영상물을 제작할 때는 실사 촬영물에 CG를 합성하는 방식으로 진행된다. 이때 실사 촬영 시 배우 위치와 행동이 추후 후반작업에서 합성될 CG 결과물의 위치와 행동이 맞지 않아 오류가 발생하게 되는데 이러한 경우는 재촬영이 필요하며, 재촬영에 대한 비율

은 업계 평균 10% 내외로 알려져 있다.

실험의 결과를 입증하기 위해 영화 명량의 러닝타임과 제작비, CG제작이 필요한 프레임 수를 기준으로 한다. 순제작비는 150억 원이고, 이 금액 중 CG 처리 비용은 43억이 소요되었으며, 부대비용은 107억 원이 소요되었다. CG가 사용된 프레임 수는 1,200프레임이며, 프레임 당 제작비용은 약 3,583,000원이 소요되지만 재촬영을 하지 않았을 경우에는 약 3,258,000원으로 325,000원 정도를 절감할 수 있으므로 총 CG 제작비용에서 3.9억 원의 절감효과를 기대할 수 있게 된다.

또한 순제작비에는 CG 처리의 중복 작업뿐만 아니라 배우 및 엑스트라, 스태프 등의 인원 동원을 비롯한 기타 부대비용이 추가로 포함되어 있다. 이러한 재촬영 비용 중에서 CG 제작비를 제외한 나머지 금액에서도 재촬영 비율에 해당하는 금액을 줄일 수 있게 된다. 재촬영은 CG를 위해 실사 촬영을 했던 부분에 대해서만 이루어지는데, 128분의 상영시간 중에서 CG 처리된 영상물의 상영시간이 50분이므로 CG 촬영 비율인 39%에 해당하는 금액 41.8억 원은 재촬영 비용이 포함된 금액이다. 따라서 재촬영 비율이 10%인 경우에는 본 시스템을 통해 재촬영을 하지 않을 수 있으므로 약 3.8억 원이 절감될 것으로 본다. 결과적으로 CG 제작비 절감액과 부대비용 절감액을 합산하여 약 7.7억 원의 절감 효과를 기대할 수 있게 된다.

## V. 결론

본 기술은 가상의 환경과 3D 모델이 삽입될 위치를 미리 영상에서 확인해 볼 수 있고, 촬영 현장에서 화면을 통해 배우의 액션과 가상 환경이 합성된 결과물을 실시간으로 모니터링 할 수 있기 때문에 포스트 프로덕션 과정에 발생하게 될 오류를 줄일 수 있게 된다. 본 시스템을 통해 재촬영 비율을 줄일 경우, 고품질 영상의 제작기간을 단축할 수 있으며, 제작비를 상당금액 절감할 수 있으므로, 다양하고 실험적인 영상물 제작의 활성화를 기대할 수 있게 된다.

이 시스템에 적용된 2D 배경은 애니메이션이나 특수효과는 적용되지 않았고, 고정된 이미지만 보여줄 수 있다. 따라서 향후 다양한 배경을 보여주면서도 애니메이션이나 특수 효과를 적용하는 기술에 대해서도

연구할 계획이다.

### References

- [1] C.-K. Lee and G. Park, "A Study on Comparison of background chroma studio for Virtual Studio," *J. of the Korea Society of Space Technology*, vol. 7, no. 2, 2012, pp. 36-41.
- [2] S. Wright, *Compositing Visual Effects*. Massachusetts : Focal Press, 2011.
- [3] H.-M. Lee, N.-H. Ryu, and E.-K. Kim, "Expression Technique of the Plurality IR LED Marker using OOK Method for Augmented Reality," *J. of the Korean Institute of Electronic Communication Sciences*, vol. 7, no. 2, 2012, pp. 433-438.
- [4] J. Park and J. Park, "Upper Body Exercise Game using a Depth Camera," *J. of Korean Society For Computer Game*, vol. 25, no. 1, 2012, pp. 61-66.
- [5] S.-M. Kim, J.-K. Song, B.-W. Yoon, and J.-S. Park, "Height Estimation using Kinect in the Indoor," *J. of the Korean Institute of Electronic Communication Sciences*, vol. 9, no. 3, 2014, pp. 343-350.
- [6] J. Webb and J. Ashley, *Beginning Kinect Programing with the Microsoft Kinect SDK*. California: O'Reilly, 2012.
- [7] K.-I. Kim, S.-H. Han, and J.-S. Park, "A Distortion Correction Method of Wide-Angle Camera Images through the Estimation and Validation of a Camera Model," *J. of the Korean Institute of Electronic Communication Sciences*, vol. 8, no. 12, 2013, pp. 1923-1932.
- [8] S.-C. Kwon, W.-Y. Kang, and Y.-H. Jeong, "Stereoscopic Video Composition with a DSLR and Depth Information by Kinect," *J. Korea Inst. Commun. Inform Sci. (KICS)*, vol. 38C, no. 10, 2013, pp. 920-927.
- [9] S.-U. Kim, "An Image Denoising Algorithm for the Mobile Phone Cameras," *J. of the Korean Institute of Electronic Communication Sciences*, vol. 9, no. 5, 2014, pp. 601-607.

### 저자 소개



#### 이혜미(Hye-Mi Lee)

2004년 순천대학교 정보통신공학과 졸업(공학사)

2010년 순천대학교 대학원 컴퓨터과학과 졸업(이학석사)

2012년 순천대학교 대학원 컴퓨터과학과 박사과정 수료

※ 관심분야 : 영상처리, 컴퓨터그래픽스, HCI



#### 류남훈(Nam-Hoon Ryu)

2009년 순천대학교 대학원 컴퓨터과학과 졸업(이학석사)

2012년 순천대학교 대학원 컴퓨터과학과 졸업(이학박사)

2012년~현재 (주)위더스비즈 대표이사

※ 관심분야 : 영상처리, 컴퓨터그래픽스, 알고리즘



#### 노관승(Gwhan-Sung Roh)

2005년 동명대학교 컴퓨터공학과 졸업(공학사)

2005 (주) 나무가 주임연구원

2009 (주) 엑시움 테크놀로지 선임 연구원

2012년~현재 온유기술 연구소 대표

※ 관심분야 : 영상처리, 컴퓨터그래픽스



#### 김응곤(Eung-Kon Kim)

1980년 조선대학교 전자공학과 졸업(공학사)

1986년 한양대학교 대학원 컴퓨터공학과 졸업(공학석사)

1992년 조선대학교 대학원 컴퓨터공학과 졸업(공학박사)

1993년~현재 순천대학교 컴퓨터공학과 교수

※ 관심분야 : 영상처리, 컴퓨터그래픽스, 멀티미디어, HCI