

종합설계 프로젝트 수행 보고서

|  |  |  |
| --- | --- | --- |
| **프로젝트명** | Unity에서의 GPGPU 기반 광선 추적 렌더러 | |
| **팀번호** | S2-2 | |
| **문서제목** | 수행계획서 | O |
| 2차발표 중간보고서 |  |
| 3차발표 중간보고서 |  |
| 최종 결과 보고서 |  |
| **팀장** | 김 한 상 | |
| **팀원** | 김 수 혁 | |
| 정 지 윤 | |
| **지도교수** | 공 기 석 | |
| 노 영 주 | |

1. 문서 정보

## **0.1 문서 수정 내역**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **작성일** | **대표작성자** | **버전** | **수정내용** | |
| 2019.01.01 | 김한상(팀장) | 1.0 | 수행 보고서 | 최초작성 |
| 2019.01.01 | 김수혁 | 1.1 | 수행 보고서 | 내용보강 |
|  |  |  |  |  |

## **0.2 문서 구성**

|  |  |  |
| --- | --- | --- |
| **진행단계** | **양식[[1]](#footnote-1)** | **내용** |
| 프로젝트 계획서 발표 | 기본 양식, 본문 양식 | 1.1 ~ 2.3 |
|  |  |  |
|  |  |  |
|  |  |  |

**이 문서는 한국산업기술대학교 컴퓨터공학부의 “종합설계” 교과목에서 프로젝트 “Unity에서의 GPGPU기반 광선 추적 렌더러”를 수행하는 S2-2 Team. SEETHELIGHTS(김한상, 김수혁, 정지윤)이 작성한 것으로 사용하기 위해서는 팀원들의 허가가 필요합니다.**

# **목차**

종합설계 프로젝트 수행 보고서 0

0. 문서 정보 1

**0.1 문서 수정 내역** 1

**0.2 문서 구성** 1

목차 2

1. 서론 3

**1.1.** **작품선정 배경 및 필요성** 3

**1.2.** **기존 연구/기술동향 분석** 7

**1.3.** **개발 목표** 8

**1.4.** **팀 역할 분담** 9

**1.5.** **개발 일정** 9

**1.6.** **개발 환경** 9

2. 본론 10

**2.1.** **개발 내용** 10

**2.2.** **문제 및 해결방안** 15

**2.3.** **시험 시나리오** 15

3. 참고문헌 10

[ 별 지 1–1 ] 17

**1.** **보고서 양식** 17

종합설계 프로젝트 수행 계획서

1. 서론

## **작품선정 배경 및 필요성**

* + 1. 배경 : General Purpose of GPU, GPU 의 범용 계산

원래 GPU는 실시간 컴퓨터 그래픽을 위하여 개발된 하드웨어이나, GPU 가 가진 다른 계산 방식의 특성을 살려 GPU 자체를 여러 프로그램을 동시에 실행시키는, GPGPU로 사용하는 기술이 2000년대에 출시되었다. CPU는 많은 캐시메모리/레지스터를 이용한 메모리 I/O 속도를 빠르게 하여 적은 수의 강력한 코어를 사용하는 strong/multi-core 전략이였다면, GPU는 많은 코어 수를 가지나 이에 대한 한계로 적은 레지스터/캐시메모리를 가져, 보다 많이 느린 메모리 I/O 속도를 가지는 week/many-core 전략을 가진 형태로 발전했다.

이는 기존 GPU 시장의 가능성을 열어주는 것으로 굉장히 획기적 이였으나, 기존의 폰 노이만 구조에 익숙해진 인력풀이 GPU를 사용하는 이기종 컴퓨팅 기술에 진입할 시간이 부족했으며, 이를 리드할 수 있는 산업의 인력 풀 자체가 크지 않았기 때문에 지금처럼 크게 알려지지는 않았었다. (많은 경우의 B2B 상품에 사용, 예:MRI) GPU 기술의 리드는 NVidia 에서 담당했는데, 기존의 수익구조가 뚜렷하고(조립 PC 시장), 연구진들에 대해 꽤 많은 투자를 꾸준하게 했기 때문에 GPU H/W, S/W 기술은 점진적으로 계속 발전하고 있었다.

2012년, AlexNet 이 인공 신경망에 대한 GPU 의 활용으로 인공 신경망 자체의 실용성, 실현 가능성을 보여주어 딥 러닝 연구들의 활성과 함께 GPU 산업 자체가 빠르게 성장하기 시작하여 현재 많이 쓰이는 딥러닝 프레임워크(tensorflow, café, ..)들은 대부분 NVidia의 cuda 기반 API 인 cuDNN 을 사용하여 구현한 것으로 시작되었다. 그림 1에서 이를 볼 수 있다.

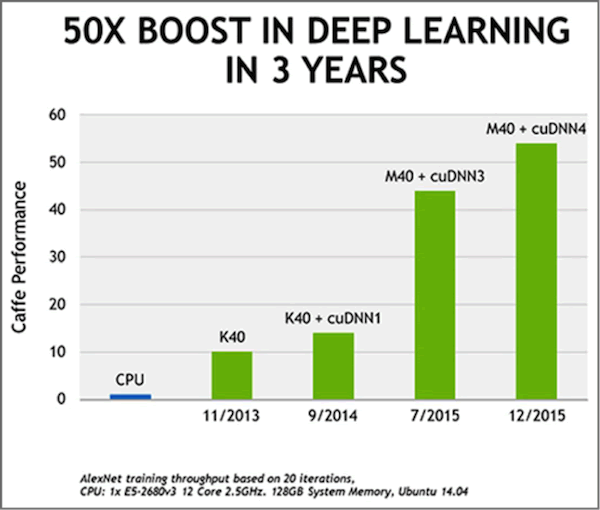


그림 : AlexNet의 CPU 및 각종 GPU & 라이브러리와의 성능 비교

* + 1. 배경 : Ray-Tracing, 광선 추적

컴퓨터 그래픽 기술은 크게 두가지, Ray-Tracing 그리고 Rasterization으로 나뉜다. 이들은 전부 3D 로 된 물체의 위치와 각 물체의 표면의 특징, 광원과의 상호작용을 계산하여, 하나의 2차원 색의 집합으로 만드는 것, 즉 모든 3D 물체의 위치에 따라 2D 이미지에 투영시키는 것은 같다. 하지만 ray-tracing 은 이미지의 픽셀의 위치와 카메라 방향을 따라서 물체의 존재여부를 추적하여, 이 정보들을 사용해 색을 계산한다. Rasterization 은 존재하는 각각의 3D 물체들을 homogenous transform을 사용하여 2차원 이미지 상에 투영하는 방법이다. 즉 ray-tracing 은 픽셀별로 모든 물체들을 탐색을 하고, rasterization은 각각의 물체를 이미지의 픽셀에 투영시키는 방법이다. 그림 2에 차이가 나타난다.

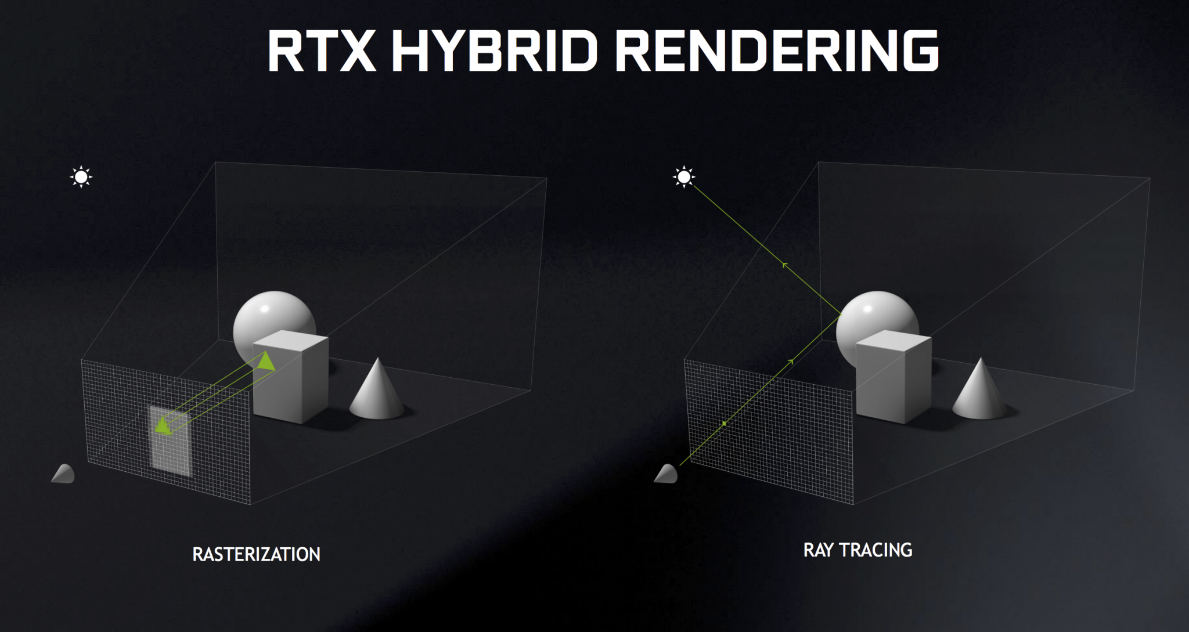


그림 : rasterization vs ray-tracing

RTX 기종이 출시되기 전, 상용 GPU들은 전부 다 rasterization 기반의 계산만을 지원했다. 기본적으로 3차원 그래픽을 게임 같은 반응성이 좋아야 하는 실시간 렌더링에 사용되기도 했고, ray-tracing은 rasterization보다 HW 가속을 지원하여 얻는 성능의 이득이 너무 적었고, 이전의 경우에는 HW 제조의 입장에서도 크게 차이가 나기도 했었기 때문에 rasterization으로만 HW 가속을 지원했었다. 그리하여 graphics API 인 OpenGL, DirectX 모두 rasterization을 위한 API 로 구성되어 있다.

앞서 언급한 딥 러닝의 도구로 GPU가 사용되면서, NVidia 에서는 기존의 프로그램만을 돌리는 CUDA Core 를 변형하여, Tensor Core 라는 이름의 딥 러닝에서 필요한 Matrix 계산을 빠르게 할 수 있는 코어를 개발하여 CUDA Core + Tensor Core 를 가진 워크스테이션용 가속기(GPU의 역할을 수행하지 못하는 병렬 연산만을 제공하는 칩)을 만들었다. 이에 영감을 얻어 다음 신제품을 발표할 때, CUDA Core + Tensor Core 와 동시에, Ray-Tracing을 계산하는 RT Core를 가진 RTX 제품을 출시했다. 즉 하나의 칩에 적어도 3가지의 큰 종류를 가진 코어들로 칩이 구성된 것이다. 동시에 DirectX API 가 이를 제어하기 위한 DirectX Ray-tracing API를(이하 DXR) 지원하기 시작했다.

하지만 HW 가속을 받는 경우에도 모든 것을 Ray-tracing으로 계산하지는 않는다. 모든 것을 계산하기 위해서는 아직 HW 자체의 지원이 부족하다. 그래서 이전부터 Rasterization으로 한계가 있는 부분들을 Ray-tracing을 모방한 기법들로 해결했던 부분을 이제는 Ray-Tracing 자체를 활용하여 그 한계를 서로 메꿔주는 방법을 사용하는 Hybrid Rendering 방식이 지금은 Ray-Tracing을 사용하는 가장 나은 방법이라고 할 수 있다.

Hybrid-Rendering 방식은 Rasterization 이 기반이기 때문에 Photorealistic한 이미지를 만들어 내기 위해서는 많은 사전처리가 필요하다. 특히 이 작품에서 ***중점***적으로 고려하는 **빛의 움직임의 결과를** 미리 계산하여 저장하는 방법을 고민하고, 이를 달라지는 3D 환경에 따라서 직접 전부다 계산을 돌려주어야 한다. 또한 성능, 결과와 관련하여 이 알고리즘들과 맞물려 조합되는 것들이 존재하기 때문에 이들은 꽤나 골치 아픈 부분이다.

만약 시간의 제한이 꽤 적은 편이고, 모든 것을 Ray-Tracing에서 심화된 Path-Tracing을 사용하여 계산할 경우, 사용자는 쉽게 사실적으로 빛이 표현된 이미지를 얻을 수 있다. 이는 게임 같은 반응성이 굉장히 중요한 실시간 렌더링에서는 사용될 수 없지만, 이외의 반응성이 필요 없는 경우에는 유용하다고 할 수 있다.

* + 1. 배경 : Unity

Unity는 UE4와 함께 일반적으로 많이 쓰이는 상용 게임 엔진 중 하나다. 아래 두개의 통계 자료에서 이를 알 수 있다. 크로스 플랫폼을 지원하며, 다른 상용 엔진에 비해 바닐라 버전이 light-weight 이기 때문에 모바일에서 많이 쓰이는 것을 알 수 있다. 그림 3에서 이를 볼 수 있다.

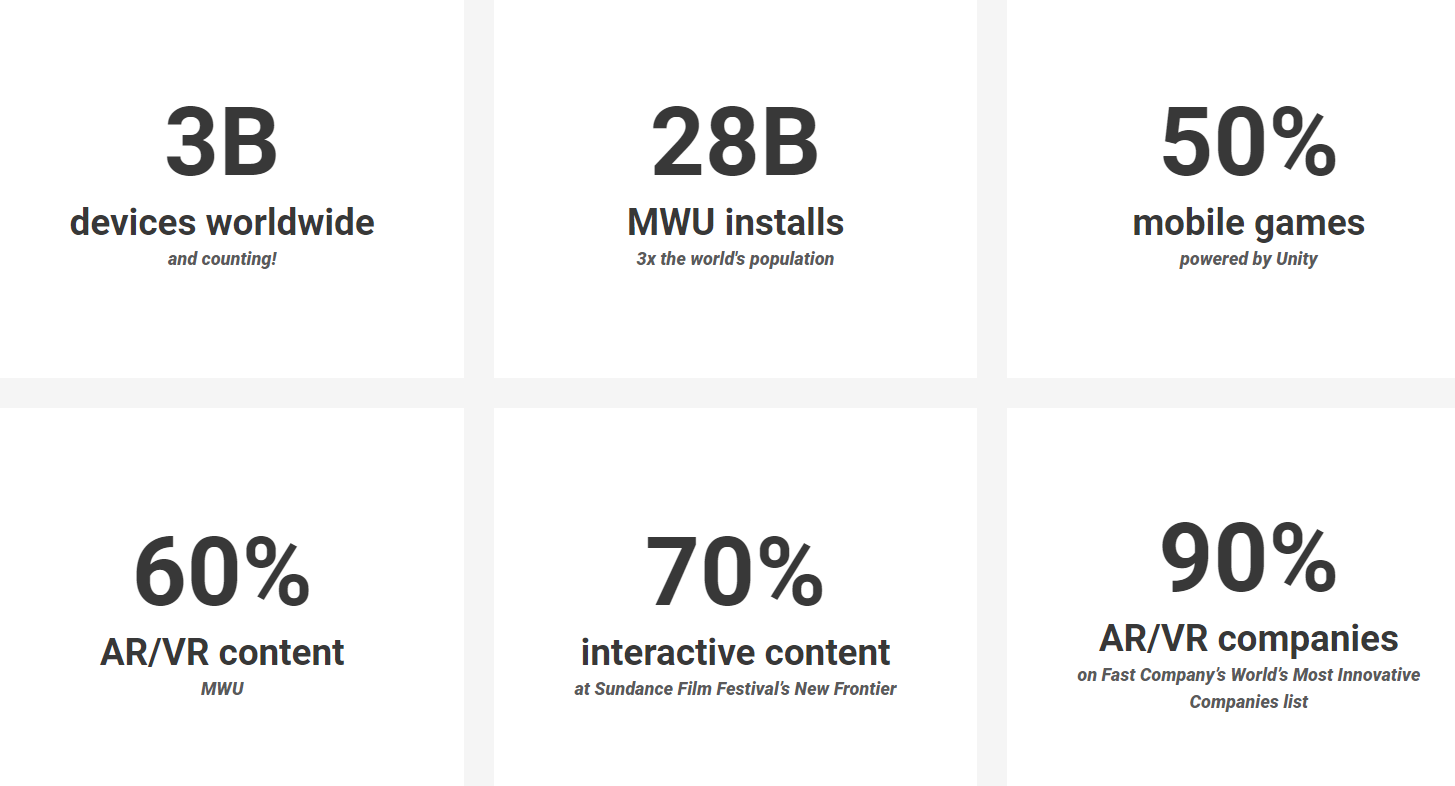


그림 : Unity 사용률 ([Powered by Unity](https://unity3d.com/public-relations))

아래 그림 4는 2014년에 진행된 설문에서 Unity 사용률이 가장 높은 비율을 차지한 것을 볼 수 있다. 100명의 응답으로 정확하진 않지만 결과적으로 많은 사용자들이 Unity를 사용한다는 것을 알 수 있는 자료로는 충분하다.,

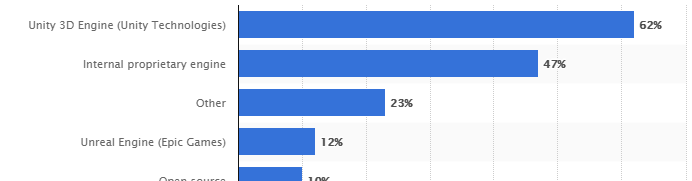


그림 : 엔진 사용 통계(영국, 2014, 100명의 응답, 비디오 게임 개발자 대상)([stastista.com](https://www.statista.com/statistics/321059/game-engines-used-by-video-game-developers-uk/))

또한 VR 디바이스의 등장으로, VR 컨텐츠 들이 주목받기 시작하며 많은 컨텐츠들이 쏟아져 나왔는데, Unity가 이에 크게 기여한 것을 위의 통계자료에서 알 수 있다. 주목해야할 사실은 이전의 주로 사용되던 출력 디바이스인 모니터에 비해 체험의 용도로 활용성이 높은 VR 디바이스에 많은 사람들이 집중되었고, 상용 엔진으로 제작하는 VR 컨텐츠는 단순히 게임만이 아니라 인터랙티브 컨텐츠로 일반화되는 계기가 되었다. 이에 더불어 상용 엔진들이 시네마틱 에디팅을 지원함으로써, 디지털 영상 컨텐츠 제작 툴로써의 길을 열게 되었다. 그림 5는 에디팅을 하는 모습을 직접 담았다.



그림 : Unity 타임라인, 시네마틱 에디팅 기능을 지원

* + 1. 필요성

실시간 렌더링 기반의 상용 게임 엔진을 사용하는 사용자 중, 영상 제작의 용도로 사용하는 사용자들이 늘어나게 되었다. 대부분은 선형 분기의 동영상을 만드므로, 이를 미리 계산해 동영상으로 만드는 기능을 원한다. 또한 컨텐츠 제작자의 필요에 따라, 사실적인 빛의 표현을 원하는 경우가 있을 수 있다. 이러한 경우, 앞서 언급한 GPGPU 기반으로 Ray-Tracing을 계산하여 나은 결과를 보여줄 수 있다.

## **기존 연구/기술동향 분석**

1.2.1 OctaneRender

Unity에서 가장 널리 알려진 Ray-Tracing을 활용한 이미지 생성기는 OctaneRender가 존재한다. 이는 Unity에서 약간의 씬 에디팅을 통해 환경을 구성하면, 이를 미리 구현된 플러그인에 넘기어 이미지를 생성한다.

Unity에서 OctaneRender의 파이프라인은 Unity에서 C#, managed memory 기반으로 3D 데이터(메쉬, 물체 별 변환(위치,회전,크기), 재질, 광원)를 자체적으로 구현한 unmanaged memory 기반의 C/C++와 CUDA를 사용한 구현체에서 실질적인 Ray-Tracing 계산을 수행한다. 이는 기본적인 파이프라인에 대한 실질적인 제시라고 할 수 있다.

Unity에서 GPGPU를 사용하기 위해서는 DirectX에서 지원하는 compute shader나, OctaneRender처럼 C#과 C/C++의 상호운용성을 이용하여 CUDA 혹은 OpenCL로 실행해야 한다. compute shader는 일반화된 API이라서 더 세밀하지 못하기 때문에, 퍼포먼스를 생각한다면 CUDA/OpenCL로 갈 수밖에 없다.

OctaneRender는 각 DCC[[2]](#footnote-2)/ 상용엔진 별로 많은 라이선스가 존재한다. Unity의 경우, 하나의 GPU에서 계산하는 경우는 무료, 두개의 GPU에서 계산하는 것부터 유료 구독형 라이센스로 전환된다. 만약, 두개의 GPU의 계산을 지원하면서 무료 라이선스를 지원한다면, 더 나은 서비스라고 할 수 있다.

또한, OctaneRender의 기능들은 실시간 렌더링에(lightmapping) 포커스가 맞춰져 있는 만큼, 영상 제작의 편의성에 있어 더 나은 기능을 보인다면, Unity 플러그인 분야의 다른 시장을 공략한다고 할 수 있다.

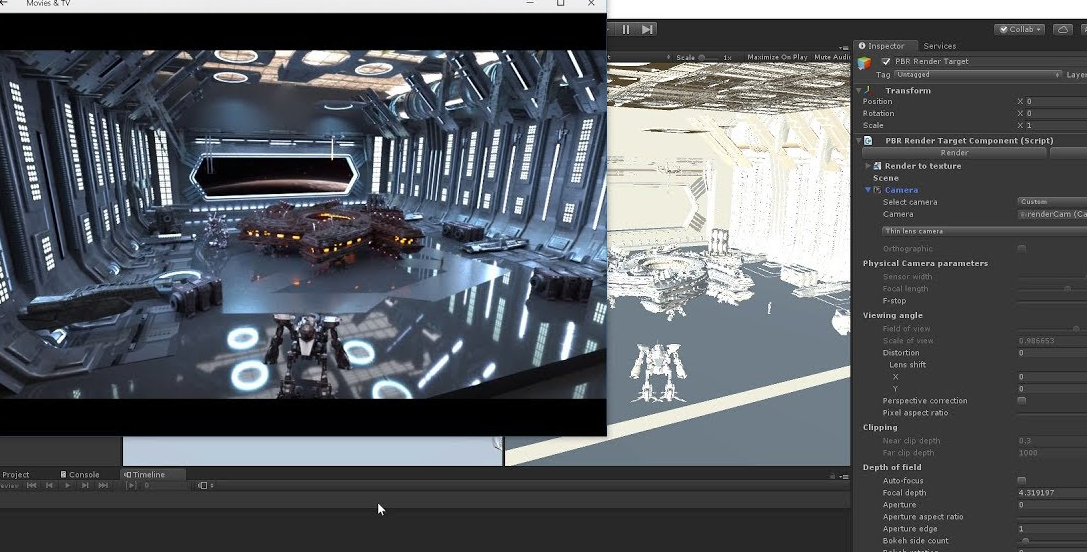


그림 : OctaneRender in Unity

1.2.2 Ray-Tracing in HDRP/DXR

RTX 시리즈에서 Ray-Tracing을 HW 자체적으로 지원하기 시작하고, DXR API가 등장함과 동시에, 이를 많은 엔진들이 차용하여 하이브리드 렌더링을 구현하기 시작했다. 이는 Unity에서 고품질의 실시간 렌더링을 지향하는 HDRP 또한 마찬가지다. 하지만 이는 렌더링의 패러다임을 바꾸는게 아닌, 단순히 기존의 Rasterization 기반 구조에서의 품질/성능 향상을 위해 부가적으로 DXR을 사용하는 것 뿐이다. 이는 현 시점의 GPU의 한계를 명확하게 보여주는 예시이다.

또한 이는 현재 Exprimental 버전이라서 안정적이지 않으며, HDRP 전용 메터리얼(재질)만 지원하므로 접근성 + 확장성에서도 한계가 명확하게 존재한다.

1.2.3 비교

|  |  |  |  |
| --- | --- | --- | --- |
|  | 목적성 | 프레임워크/메터리얼 | H/W |
| OctaneRender | 실시간 렌더링 보조  레이 트레이싱 렌더 | Standard 메터리얼 | NVidia GPU |
| HDRP/DXR | 레이 트레이싱 렌더 | HDRP 런타임/메터리얼 | DXR 지원 H/W |
| ***RadianceGrabber*** | 레이 트레이싱 렌더  영상 제작 | Standard, URP 메터리얼 | NVidia GPU |

표 : OctaneRender, HDRP/DXR, RadianceGrabber(해당 작품) 비교

## **개발 목표**

기존의 존재하는 상용엔진/DCC에서 돌아가는 사실적인 명암을 가진 이미지를 생성하는 것, 이를 연속적으로 생성한 것을 바탕으로 동영상을 생성하는 것을 기능적인 목표로 가진다.

여기에 더불어, OctaneRender와의 차별성을 두기 위해서, Multi-GPU를 사용하여 성능 상의 이점을 가지며 계산하는 기능, 보다 현실적인 방법으로는 2개의 GPU를 사용시 150%~170%의 성능을(시간 단축) 낼 수 있는 기능을 성능상의 목표로 가진다.

## **팀 역할 분담**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 김수혁 | 김한상 | 정지윤 |
| 자료수집 | 광선추적, CUDA | CUDA | Unity 렌더링 환경 |
| 설계 | 광선 추적 계산 모델 설계  (알고리즘 선택) | CUDA 기반의  Ray Tracing 계산 설계 | Unity 환경 데이터 수집 |
| 구현 | host/device 각각의 Foundation 구현 및  CUDA 기반 광선추적 구현 | | Unity Plugin의 형태로 환경 데이터 수집 및 UI |
| 테스트 | CPU/GPU 성능 별, 씬 복잡도에 따른 성능 테스트 | | |

표 : 팀원 별 역할 분담

역할 분담은 최대한 쉽게 나눌 수 있는, 각자 작업의 의존성이 적은 것들로 나누었다. 그렇지만 Unity 렌더링 환경의 세팅과, 광선 추적 모델이 먼저 어느 정도의 결과가 나온 후 구현에 들어가야 하는 부분에 있어서는 의존성이 존재한다.

## **개발 일정**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 12월 | 1월 | 2월 | 3월 | 4월 | 5월 | 6월 | 7~9월 |
| 조사 및 학습 |  | |  |  |  |  |  |  |
|  | |
|  | |
| 계산 모델 설계  SW 설계 |  | |  |  |  |  |  |  |
|  | |
|  | |
| 구현 |  |  | | | |  |  |  |
|  | | | |
|  | | | |
| 테스트 및 데모 |  |  |  |  | | |  |  |
|  | | |
|  | | |
| 문서화 및 발표 |  |  |  |  | | | |  |
|  | | | |
|  | | | |
| 최종보고서  작성 및 발표 |  |  |  |  |  |  |  |  |
|  |
|  |

표 : 개발 일정, 2019.12 ~ 2020.09

## **개발 환경**

개발 언어/환경 C++, CUDA / Visual Studio 2017, NVidia CUDA Toolkit 10.1 Update 2

C# / Unity 2019

개발 기준 HW PC1 I7-4790 / DDR3 16GB / GTX 970

PC2 I7-6700 / DDR4 32GB / GTX-1070

# **본론**

## **개발 내용**

개발에 필요한 노력을 크게 나누자면, 2가지로 나눌 수 있다.

첫번째는 Unity에서 기능을 제공하기 위해 직접 C#을 이용하여 작성한 일종의 Front-End 모듈이다. 가장 중요한 기능은 C++/CUDA로 작성된 광선 추적 모듈에 넘겨야 할 데이터들을 제공해주는 부분인데, 이는 Unity에서 노출된 C# 기반의 데이터를 광선 추적 모듈의 정의된 형식, ABI에 따라서 넘겨준다. 광선 추적 모듈에 넘기는 정의된 데이터의 종류는 총 8가지이다.[[3]](#footnote-3) 아래 표 3에서 이를 볼 수 있다.

|  |  |  |  |
| --- | --- | --- | --- |
| 데이터 | 세부 데이터 | | 설명 |
| 카메라 | 위치, 회전, 정의된 투영 행렬 | | 3D 공간에서 물체를 그리기 위한 기준의 주체 |
| 스카이박스 인덱스 | |
| 스카이박스 | Cubemap | 육면체 | 아무런 물체가 그려지지 않을 때, 기본으로 그려지는 배경 |
| Procedural | 임의의 Shading |
| Paranomic | 구의 표면을 펴놓은 사각형 |
| 광원 | Directional | 하나의 방향의 광원/배경 | 빛의 원인,  모양에 따라서 광선의 추적 시, 판정이 달라짐 |
| Point | 구 형태의 광원 |
| Spot | 원뿔 형태의 광원 |
| Area | 표면(주로 사각형)형태의 광원 |
| 메쉬 | 정점 버퍼 데이터 | 3차원상의 위치, 법선, 텍스쳐 접근을 위한 재매개화된 좌표 등의 데이터를 저장.  Graphics API에서 제공 | 하나의 물체를 나타내는 기하학적 정보들의 집합. |
| 인덱스 버퍼 데이터 | 정점 버퍼의 인덱스로 구성되어, 하나의 면(삼각형, 사각형, ..)을 이루게 하는 정보  Graphics API에서 제공 |
| Bindpose 행렬 데이터 | 초기 세팅된 위치에서 “뼈”의 위치, 회전 정보를 없에주는 행렬, 각 “뼈” 별로 저장함. |
| 메쉬 렌더러 | 메쉬 참조 인덱스 | 참조할 메쉬 인덱스, 하나만 존재 | 메쉬와 메터리얼을 참조하고, 위치,회전,크기 정보들을 가짐. |
| 메터리얼 참조 인덱스 | 참조할 메터리얼 인덱스,  여러 개가 존재 할 수 있음. |
| 동차 변환 행렬 | 위치, 회전, 크기 정보를 가지는  변환 행렬 |
| AABB | 메쉬를 감싸는 박스 형태의 모양. |
| 스킨메쉬 렌더러 | 메쉬 참조 인덱스 | 참조할 메쉬 인덱스, 하나만 존재 | 메쉬 렌더러의 데이터와 더불어, “뼈”들의 위치, 회전 정보를 가지고 있음. |
| 메터리얼 참조 인덱스 | 참조할 메터리얼 인덱스,  여러 개가 존재 할 수 있음. |
| 동차 변환 행렬 | 위치, 회전, 크기 정보를 가지는  변환 행렬 |
| AABB | 메쉬를 감싸는 박스 형태의 모양. |
| “뼈”의  위치, 회전 | 각 “뼈”의 위치와 회전 정보  스키닝 계산시 사용됨. |
| 텍스쳐 | 1D | 1차원 좌표로 접근 | 데이터를 1D,2D,3D로 저장하여 샘플링 할 수 있는 형태의 GPU 특유의 데이터 구조 |
| 2D | 2차원 좌표로 접근(주로 쓰임) |
| 3D | 3차원 좌표로 접근 |
| 메터리얼 | 쉐이더 | Unity에서 접근하는 프로그래밍 가능한 코드,  여기서는 Unity의 기본 Shader만 지원함. | 쉐이더와 그 쉐이더에 넘길 파라미터들을 가지고 있는 인스턴스 |
| 파라미터 | 해당 Shader에서 정의한 매개변수  고정된 쉐이더만 지원하기 때문에 이를 하드 코딩으로 지원. |

표 : 입력 데이터 명세

표 3의 정보들을 Unity 에디터 상에서 C#을 통해 자체적으로 Serialization을 한 다음, C++에 넘겨준다. 그리고 여러 개의 프레임을 한꺼번에 계산할 때가 아니라, Unity에서 환경을 구성하면서 Ray-Tracing 결과를 미리 보고 싶은 경우 한 프레임을 점진적으로 보여주어야 한다. 이때는 결과를 저장할 2D 텍스쳐를 생성한 후, 텍스쳐 포인터와, 함수 포인터를 넘겨주어 광선 추적 모듈에서 이를 호출해 다시 텍스쳐를 그리거나, 에디터상에서 주기적으로 다시 그린다.

쉐이더의 경우, 해당 쉐이더의 픽셀 쉐이더 함수의 구현에 따라서 Ray-Tracing에서 구현할 것 또한 달라진다. 그래서 모든 쉐이더에 맞추어 구현을 할 수 없고, 현실적인 제약 조건으로는 Unity에서 제공하는 Standard, Standard(Specular Setup), URP(LWRP)의 Lit 쉐이더를 Ray-Tracing과 연관시켜 구현할 계획이다. Unity에서의 환경 조성 시 Standard가 기본이기 때문에, 컨텐츠 제작의 용도로 사용되기 적합할 것으로 보인다.

영상 녹화 기능의 경우, OpenCL에서 C#형태의 라이브러리를 지원하고, OpenCL 자체에서 영상 녹화 기능을 프레임 단위로 제공하기 때문에, 이를 사용하여 영상 녹화 기능을 제공한다.

에디터 상에서 필요한 UI는 Unity에서 제공하는 IMGUI 라이브러리를 사용하여 구현한다. 현실적으로 가능한 것은, 간단하고 직관적으로 영상 녹화 기능을 제공해야 한다.

두번째 노력은 광선 추적기 구현에 앞선 알고리즘 선택 및 설계와 이에 대한 구현의 반복이다. 알고리즘 선택 및 설계에서 고려해야할 것은 3가지가 있다. 첫번째로는 공간상의 물체들을 빠르게 추적하기 위한 자료구조를 선택하는 것이다.

일반적으로 Ray-Tracing에 사용되는 자료구조는 Bounding Volume Hierarchy(이하 BVH)라는 방법으로, 공간상의 물체를 특정 축을 기준으로 정렬한 후, 이를 트리의 anary에 따라서 묶어서 최종적으로 모두 이어진 트리 한 개를 가지게 되는 자료구조이다. Ray-Tracing을 기준으로 construction, traversal 알고리즘이 많이 연구되어 이에 대한 것들을 사전 조사 후 구현해야 한다. GPGPU를 기준으로 연구된 것도 꽤 존재하여, 도움을 얻을 수 있을 것으로 예상된다.

다음으로 연구해야 할 알고리즘은 Path-Guiding 알고리즘이다. Ray-Tracing의 고 비용, 고 품질 방법인 Path-Tracing 방법 중 가장 고전적이고 naïve한 방법이나 쉽게 구현할 수 있는, 무작위로 방향을 정하여 광선을 쏘아 해당 표면으로 이동하여 이를 반복하고, 빛을 발견하면 해당 색을 recursive하게 계산하는 방법은 치명적인 단점이 있다.

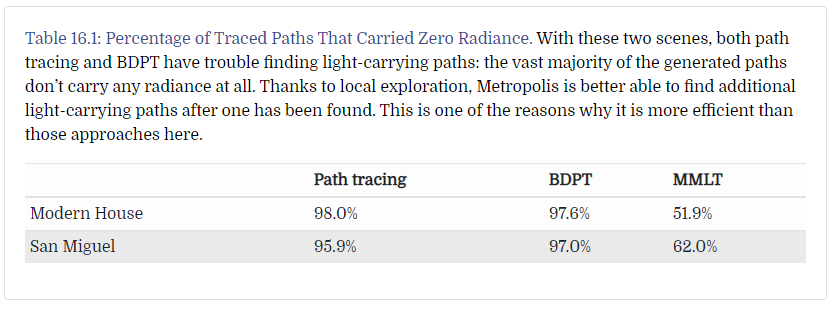


그림 : 0의 에너지를 발견한 경우, PT vs BDPT vs MMLT ([pbr-book.org](http://www.pbr-book.org/3ed-2018/Light_Transport_III_Bidirectional_Methods/Metropolis_Light_Transport.html))

그림 8에서 볼 수 있다시피, 광원을 발견하지 못해 luminance가 0인 전혀 필요 없는 샘플링이 Path Tracing, BDPT는 대부분을 이룬다. 전 문단에서 언급한 방법이 Path Tracing에 해당한다. 옆의 MMLT라는 것과 비교하면 현저히 차이가 나는 것을 알 수 있다.

즉, 유효한 빛의 경로로 샘플링을 하게 하는 것, 즉 Path-Guiding이 Path-Tracing에 있어 중요한 방법으로 떠오르게 된다. 여러 방법 중 가장 실효성이 있고 유명한 방법이 Metropolis Light Transport(이하 MLT)이다. MLT는 1997년에 발표된 기법으로, 이 이후에도 많은 변종들이 만들어졌다. 이에 이름이 알려진 것들은 PSSMLT, MMLT가 존재한다.

이러한 Path-Guiding방법은 샘플링 비용을 현저하게 줄일 수 있어, 적은 개수의 샘플링 비용 제약이 존재할 시 noise 제거와 이미지에서의 빛의 전반적인 질이 올라간다. 그림 9에서 이를 볼 수 있다.

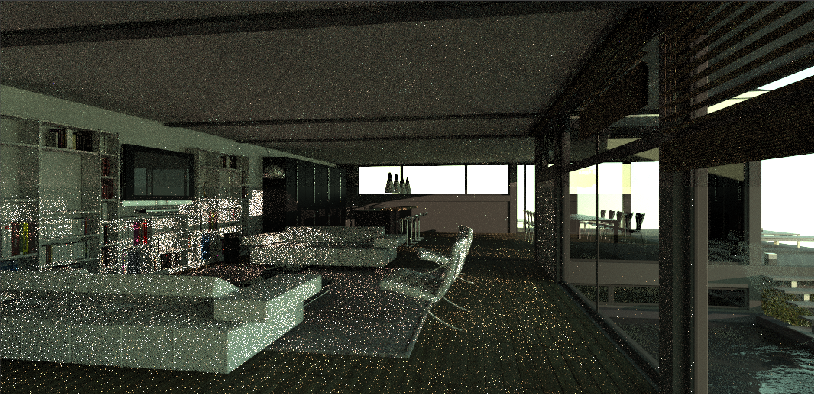


그림 : MMLT(위 절반), Path-Tracing(아래 절반), 같은 시간의 계산결과 ([pbr-book.org](http://www.pbr-book.org/3ed-2018/Light_Transport_III_Bidirectional_Methods/Metropolis_Light_Transport.html))

마지막으로는 noise 제거 알고리즘의 연구이다. MLT는 이를 최대한 막기 위한 방법으로 알려졌지만 근본적인 해결책은 아니다. 앞서 말한 것들 보다는 덜한 노력이 필요하지만 방법의 연구는 반드시 필요하다.

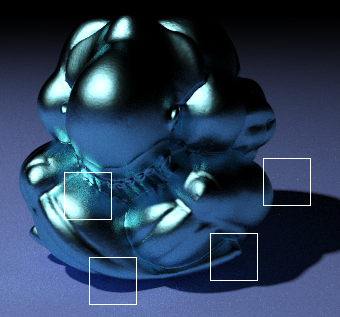
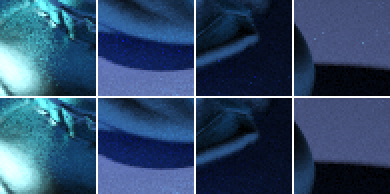
 

그림 : PT의 결과, noise(좌측), 확대한 이미지(우측/위), variance reduction한 결과(우측/아래)

다만, 그림 10에서 보이는 것처럼 이들은 흔히 알려진 post-processing(cross bilateral filter, …) 방법을 통하여 어느 정도 제거될 수 있을 것으로 보인다.

구현은 unmanaged memory상에서 CPU와 GPU를 활용하기 위하여, C++과 CUDA를 함께 사용하여 구현해야 한다. 다만 구현에 있어 퍼포먼스가 가장 중요한 요소이기 때문에, OOP를 크게 활용하지 않을 가능성이 높다.

미리 정의된 Graphics API의 데이터를 받아오기 때문에 이들을 처리하기 위해서는 DirectX, OpenGL과 같은 라이브러리로 처리를 직접 해주어야 한다. CUDA에서는 이를 처리하는 방법들을 지원하기 때문에 CPU레벨에서 계산하지 않을 시 직접 메모리 공간을 처리해주지 않아도 된다.

순서는 기존에 Unity/C# 레벨에서 넘겨준 데이터들을 버퍼를 할당하여 복사한 후, 먼저 메시를 처리해 주어야 한다. 앞서 언급한 스킨 메시/스킨 메시 렌더러의 경우, 각 정점의 위치를 “뼈”의 가중치와 인덱스를 활용하여 존재하는 bindpose matrix를 가중합한 행렬을 통해 기존 정점의 위치에 저장된 “뼈”와 관련된 위치와 회전을 제거해주고, 그 다음 각 “뼈”의 현재 위치와 회전을 적용하여, 정점의 위치를 재계산한다. 이 과정에서 CUDA를 활용하여 계산해야 한다. 원래의 경우에도 GPGPU로 계산하기 때문에 기존의 실시간 렌더링에서 하는 것과 동치이다. 테셀레이션은 Unity에서 제공하는 Standard, URP(LWRP)에서 지원하지 않기 때문에 초기 버전에서는 제외한다. 만약 구현한다면 지금 언급한 스키닝과 같은 타이밍에 처리된다.

다음은 Path-Tracing에 앞서 BVH를 생성해야 한다. BVH는 물체들을 트리의 형태로 저장하여, 일반적으로 O(n)의 시간 복잡도를 가지는 것을, O(logn)의 시간을 가지도록 하는 오브젝트에 따라 트리를 구성하는 방법이다. 이 부분은 많은 자유도가 존재하기에, 하나의 정확한 타협점을 찾기 위해서는 여러 방법, 여러 환경에 따른 많은 프로파일링이 필요하다.

하지만 데이터가 정해져 있으므로 그 부분에 맞춰야 할 필요가 있다. 일반적으로 여러 논문에서는 BVH 구성 시 리프 노드에는 폴리곤 (인덱스 버퍼의 정보로 구성된 면(Plane))의 참조 정보를 넣어 놓는다. 하지만 여기에서는 조금 다른 방식으로 구성된다. 각 물체들의 위치, 회전 정보와, 메쉬는 같이 구성되어 있지 않고, 참조 형식으로 분리되어 있다. 그렇다면, 오브젝트들의 위치, 회전을 저장한 것과, 각 메쉬들을 따로 BVH로 구성하여 2 Layer 구조를 가지게 하는 방법이 가장 처리를 덜하는 방법이다.

하지만 위와 같은 방법은 충분히 달라질 수 있다. BVH 생성시에는 꽤나 많은 코스트가 들지만, 그 뒤의 광선을 가지고 BVH를 차례차례 타고 들어가는 Ray-Traversal의 성능이 가장 중요하기 때문이다. 어디까지나 이는 구현 반복의 1단계라고 말할 수 있다.

메쉬의 개수가 굉장히 많다면, VRAM ~ RAM 간의 메모리 통신을 하지 않고 GPGPU를 통해 처리하는 것도 하나의 방법이다. 하지만 대부분 적은 경우가 많기에 CPU에서 처리하는 것이 더 빠를 것으로 예상된다.

Path-Tracing 계산 방법은 크게 두가지로 나뉜다. 하나는 무작위로 경로를 찾는 방법, 하나는 MLT를 사용하여 더 high-contribution인 경우에 더 높은 확률로 경로를 결정하는 방법이다. MLT를 사용하는 방법은 연구가 필요하기에 맨 처음에는 단순한 무작위 경로 찾기부터 구현한다. 또한 레퍼런스 이미지 생성에는 완전한 균일한 샘플링 방식으로 수많은 샘플링 횟수를 들여야 하기 때문에 MLT의 평가를 위해 필요하다.

위의 언급한 반복의 의미는, 기본적인 틀을 위해 한번 위의 언급한 순서대로 쭉 구현한 다음, 변경점이 생길 시에 부분부분 바꾸어 가면서 구현하는 것을 의미한다. BVH의 구성의 경우도 맨 처음에는 데이터에 맞추어 구성하나, MLT의 구현 이후 시간이 허락한다면, 여러 논문의 기법들을 참고하여 다양한 방법을 찾을 수 있다.

Path-Tracing 구현

BVH 구축/탐색 구현

Unity Wrapper 구현

MLT 변종 구현

Mesh Skinning 구현

MLT 구현

BVH 변종 구현

그림 : 구현 순서도

실질적인 구현의 순서는 다음과 같다. 최소한의 틀을 구현하는 것이 MLT 구현 까지고, 그 다음 부터는 추가적인 구현으로 점선으로 표기하였다.[[4]](#footnote-4)

마지막으로 이 SW에서 중요하게 판단되어야 할 것은, 비교적 짧은 시간 내에, 사람의 눈으로 인식하여 만족할 만한 빛의 시뮬레이션을 나타내는 것이다. 이를 엄밀하게 판단하기 위해서는 위 요소들을 메트릭으로 나타내야 할 필요가 있다.

첫번째 요소는 생성 시간을 측정하여 나타낸다. 이는 다음에 언급할 두번째 요소보다 굉장히 간단하다. 두번째 요소인 사람의 눈으로 인식하여 만족할 만한 것은 균일하게 랜덤으로 경로를 찾는 Path-tracing 방법으로 수많은 샘플링을 들여 만든 이미지(Reference Image, Ground Truth)와 현실적으로 가능한 샘플링 수를 적용하여 만든 이미지(Synthesized Image)를 IQA[[5]](#footnote-5) 기법으로 메트릭을 계산한다. (Joss, Mark, & Rafal, 2017) 에 따르면, 수많은 IQA 기법 중 MS-SSIM 혹은 SC-QI 기법의 사용을 추천한다. 이는 Reference Image의 노이즈를 고려해보았을 때, 가장 나은 기법이라고 말한다.[[6]](#footnote-6)

## **문제 및 해결방안**

## **시험 시나리오**

기본적인 테스트 파이프라인은 단순하다.

1. Unity에서 테스트에 필요한 씬을 세팅한다.
2. 무작위로 탐색하는 Path-Tracing으로 몇 만개의 샘플링 횟수를 사용하여 레퍼런스 이미지를 생성한다.
3. 테스트할 기법을 선택하여 세팅한 후, 현실적으로 계산할 수 있는 샘플링 수만큼 설정 후 Path-Tracing을 실행한다.
4. 이미지 생성 시간과 레퍼런스 이미지, 생성된 이미지를 IQA 기법으로 계산하여 결과를 도출한다.

# **참고문헌**

AllenRoger. (2018). “Accelerated Ray Tracing in One Weekend in CUDA”. NVIDIA Developer: https://devblogs.nvidia.com/accelerated-ray-tracing-cuda/에서 검색됨

Hachisuka, T., Kaplanyan, A. S., & Dachsbacher, C. (2014). *Multiplexed Metropolis Light Transport.*

HainesEric, & Akenine-MöllerTomas. (2019). “Ray Tracing Gems: High-Quality and Real-Time Rendering with DXR and Other APIs.” Apress.

Joss, W., Mark, J. W., & Rafal, M. (2017). Analysis of reported error in Monte Carlo rendered images. *The Visual Computer*, 705-713.

Keleman, C., & Szirmay-Kalos, L. (2002). *Simple and Robust Mutation Strategy for Metropolis Light Transport Algorithm.*

Kirk, D. B., & Hwu, W.-m. W. (2010). *대규모 병렬 프로세서 프로그래밍.* (순. 하, 크. 김, & 영. 이, Trans.) BJ 퍼블릭.

Lafortune, E. P., & Willems, Y. D. (1998). *Bi-directional Path Tracing.*

Lauterbach, C., Garland, M., Sengupta, S., & Leubke, D. P. (2009). Fast BVH Construction on GPUs. *EuroGraphics.*

Lin, D., Shkurko, K., Mallett, I., & Yuksel, C. (2019). Dual-Split Trees. *Symposium on Interactive 3D Graphics and Games.*

Pharr, M., Jakob, W., & Humphreys, G. (2016). *Physically Based Rendering.* Morgan Kaufmann.

Sanders, J., & Kandrot, E. (2010). *CUDA By Example, NVidia.* Addison-Wesley.

ShirleyPeter. (2018). “Ray Tracing in One Weekend.”

ShirleyPeter. (2018). “Ray Tracing in The Next Weekend.”

ShirleyPeter. (2018). “Ray Tracing in The Rest of Your Life.”

Veach, E., & Guibas, L. J. (1970). *Metropolis Light Transport.*

Ylitie, H., Karras, T., & Laine, S. (2017). Efficient Incoherent Ray Traversal on GPUs Through Compressed Wide BVHs. *High-Performance Graphics.*

# [ 별 지 1–1 ]

## **보고서 양식**

* 1. 기본 양식

글씨체 – 나눔고딕, 10p

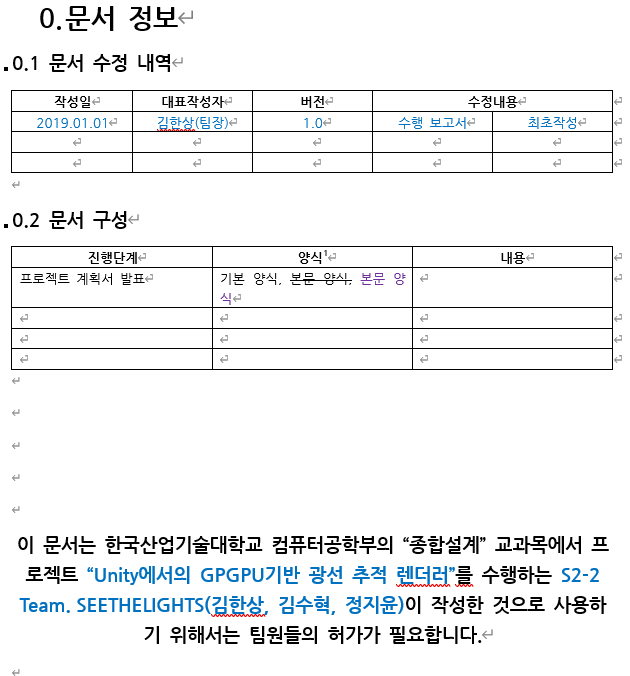
바닥글 – 페이지 번호, 아래쪽, 두줄 2

필요 시 각주 삽입

인용 – 스타일 APA, Word의 출처 관리 기능을 통해 기록 후 각 문서 마지막 부분에 추가

* 1. 문서 정보 양식

문서 정보 작성 예시



**제목 – 20p 나눔고딕 굵게**

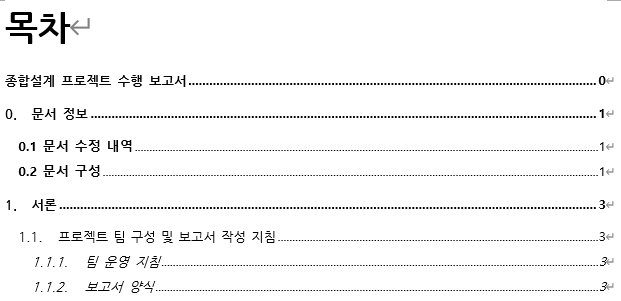
**중제목 – 14p 나눔고딕 굵게**

표 – 가운데 정렬

수정사항 자주색, 중요사항 빨간색, 최초작성 하늘색 (문서 수정 내역 한정)

* 1. 목차 작성 지침

목차 작성 예시



**목차 제목 – 24p 굵게**

**목차 항목 – 10p 굵게, 3단계까지 표시**

Word의 참조 – 목차 기능을 통해 추가하고 텍스트 추가를 통해 추가, 목차 업데이트를 통해 목차를 갱신

* 1. 본문 양식

본문 작성 예시



1. [[ 별지 1-1 ]](#_[_별_지) 참조 [↑](#footnote-ref-1)
2. “Digital Content Creation”의 약자, 여기서는 Maya, 3DS MAX 같은 3D 에디팅툴을 일컫는다. [↑](#footnote-ref-2)
3. 제안서 발표 코멘트 : “입력데이터의 scope를 잘 정의 할 것 구체화 필요” [↑](#footnote-ref-3)
4. 제안서 발표 코멘트 : “작더라도 확실한 성과가 나올 수 있도록 할 것” [↑](#footnote-ref-4)
5. Image Quality Assessment, 이미지의 품질 자체를 평가하는 방법 [↑](#footnote-ref-5)
6. 제안서 발표 코멘트 : “목표치에 대한 검증 방법 완성도가 나와야함” [↑](#footnote-ref-6)