

콘텐츠와 사용자 간 입출력 상호작용 인터페이스

정상권 IEEE 3079.2 TG 의장, (주)조이편 대표이사

1. 머리말

실감형 혼합현실 콘텐츠의 핵심은 '실감형'이다. '실감(實感)'이라는 말의 사전적 의미는 사람이 느낄 수 있다는 뜻인데, '실감형 콘텐츠'에서의 실감은 사용자와 상호작용(Interaction)을 일으킨다는 의미로 사용된다. 따라서 '실감형 혼합현실 콘텐츠'는 사용자와의 상호작용 방법으로 혼합현실 기술을 사용한다는 뜻이라고 할 수 있다.

그러한 측면에서, '콘텐츠와 사용자 간 입출력 상호작용 인터페이스'는 '실감형 혼합현실 콘텐츠'를 특정 짓는 매우 중요한 요소다.

본 고에서는 2020년 개정된 '콘텐츠와 사용자 간 입출력 상호작용 인터페이스' 표준과, 해당 표준과 연관된 표준들을 함께 살펴봄으로써 '실감형 혼합현실 콘텐츠'에 대한 전반적인 이해를 돕고자 한다.

2. 실감형 혼합현실

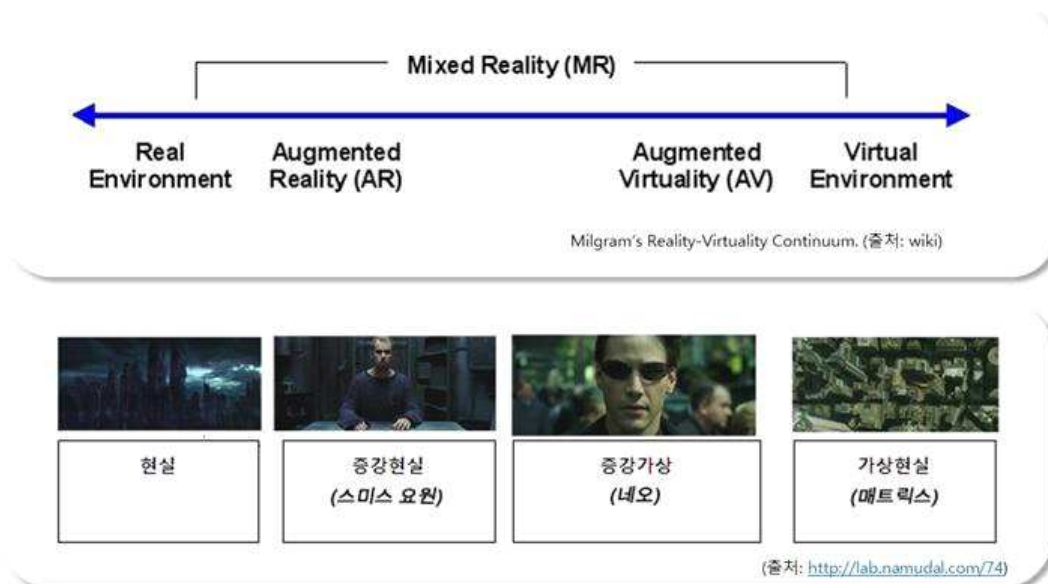
2.1 실감형 혼합현실의 구현

실감형 혼합현실은 기본적으로 사람의 동작을 읽어내는 데서 시작한다. 그래야 사람과 상호작용할 수 있기 때문이다. 사람의 동작은 키보드를 누르거나, 조이스틱을 사용하거나, 또는 터치패드를 쓰는 다양한 방법으로 컴퓨터에게 전달할 수 있다. 여기서 더 발전된 기술이라면 햅틱 또는 모션 컨트롤 정도를 생각할 수 있다.



[그림 1] 체감형 모션게임 Xtion 화면[1]

최근 가상현실(VR, Virtual Reality) HMD(Head Mounted Display) 장치들은 고차원적인 입력을 지원하기 위해 대부분 자이로 센서와 가속도 센서가 들어있는 조이스틱을 사용한다. 그러나 실감형 혼합현실 분야에서는 모션 센서에 의한 모션 컨트롤러를 많이 사용한다. 특히 룸 스케일(Room scale) 이상의 공간에서 사용자의 모션을 센싱하기 위해 적외선과 라이다(LiDAR) 센서 등을 이용한 사용자 동작 인식 방법을 많이 사용한다.



[그림 2] 혼합현실 스펙트럼의 개념[2]

혼합현실을 구현하려면 가상의 객체(virtual object)와 실제 객체(physical object)를 자연스럽게 조화시키고, 상호 간 위치정보를 정확하게 결합(synchronizing)시켜야 한다. 만약 상호 위치정보가 일치하지 않으면 사람의 동작에 대한 반응(interaction)이 일어나지 않는다.

즉, 실감형 혼합현실을 구현하려면 사람의 동작을 정확하게 읽고 분석해야 하고, 실제 객체에 해당하는 사람의 동작 좌표와 가상 객체가 위치하는 좌표의 유기적 연동(synchronization)이 필수이며, 이것이 핵심기술이다. 구글의 AR 코어도, 애플의 AR 키트도 모두 이 기술을 이질감 없이 조화롭게 구현하는 것을 목표로 한 솔루션 기술이다.

2.2 프로젝션 매핑

혼합현실이란 가상 객체와 실제 객체가 자연스럽게 상호작용함으로써 구현되는 콘텐츠 기술을 말한다. 증강현실은 그중 실제 객체를 바탕으로 가상 객체를 증강시켜 표현하는 기술이고, 반대로 가상의 객체를 바탕으로 실제의 객체를 증강시켜 표현하는 기술을 증강가상이라고 한다. 프로젝션 매핑은 실제 객체에 프로젝션으로 만들어진 가상의 객체 이미지를 투영함으로써 혼합현실을 구성하는 기법이다. 실감형 혼합현실의 기본 기술인 '가상 객체가 형성되는 좌표계와 실제 객체 좌표 간 유기적 연동'이 여기서도 콘텐츠의 핵심이 된다. 이것이 바로 '콘텐츠와 사용자 간 입출력 상호작용'이며, 이 상호작용을 위한 인터페이스가 프로젝션 매핑 기술로 구현되는 '명령어 전달을 위한 인터페이스'인 것이다.

3. 실감형 혼합현실 기술 표준

3.1 콘텐츠와 사용자 간 입출력 상호작용 지침

‘콘텐츠와 사용자 간 입출력 상호작용 지침’ 표준은 2015년 6월 10일 당시 모바일콘텐츠표준화 포럼에서 포럼표준(MCSF_02.0017)으로 제정된 이후, 12월 16일 TTA 단체 표준(TTAK.KO-10.0869)으로도 제정됐다. 당시에는 ‘혼합현실’이라고 하는 기술 분야가 가상현실에 비하여 대중의 주목을 받지 못했고 컴퓨팅 파워도 많이 부족했기에, 현재에 비해 열악한 환경을 극복해야 하는 어려움이 있었다.

동작인식 센서와 관련된 부분도 리얼센스, 키넥트 v2, 립 모션 등이 뎁스 정보를 추출하는 가장 대중적인 방법이었다. 이러한 상황에서 사용자의 동작을 완벽하게 인식하는 ‘실감형 혼합현실’을 구현하는 것은 무모한 시도였을 수 있다. 대중적으로 가장 많이 사용되는 마이크로소프트 키넥트의 동작인식 정확도(accuracy)는 키넥트 v2 기준 공식적으로 최대 70%에 불과하다.

<표 1> 키넥트 사양 비교[3]

구분		Kinect v1	Kinect v2	Azure Kinect
색상	해상도	640 × 480	1920 × 1080	3840 × 2160
	fps	30	30	30
심도	해상도	320 × 240	512 × 424	640 × 576
	fps	30	30	30
인물		6	6	5
인물 자세		2	6	5
관절		20 관절 / 사람	25 관절 / 사람	32 관절 / 사람
심도의 취득 범위		08. ~ 4.0m	05. ~ 8.0m	0.25 ~ 5.46m

이러한 상황에서 사용자의 동작을 정확도가 충분하지 않은 센서로 인식하여 명령어를 전달하는 행위의 좌표점인 ‘인터페이스와 상호작용’ 행위는 매우 정교하고 정확하게 이루어져야 한다. 이 관점에서 정의된 것이 ‘콘텐츠와 사용자 간 입출력 상호작용 지침’ 표준이다.

본래 모바일콘텐츠표준화포럼에서 제정된 포럼표준은 TTA에서 제정된 단체표준의 5.2항만으로 이루어져 있었다. 5.1항은 ‘프로젝션 기반 콘텐츠의 가상 사물의 배치 최적화 기술’, 5.3항은 ‘프로젝션 기반 콘텐츠 사용자 디스플레이 표현 최적화 기술’로 각각 모바일콘텐츠표준의 독립된 포럼 표준이었다. 이것이 TTA 단체 표준으로 개발되면서 세 포럼 표준이 하나로 합쳐지게 된 것이다.

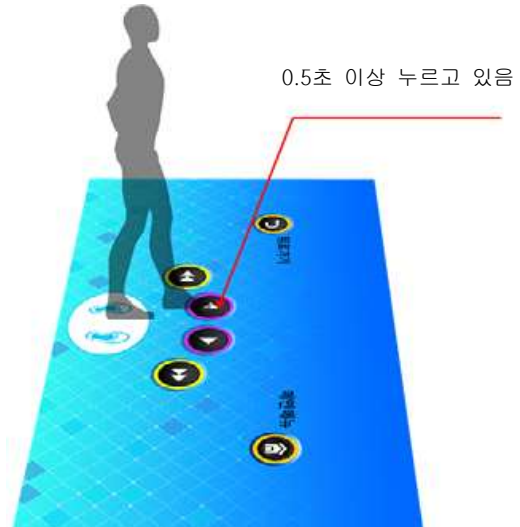
이 표준은 앞서 이야기한 바와 같이 ‘프로젝션 매핑’이라고 하는 기술을 실감형 혼합현실 콘텐츠에 적용할 때 반드시 지켜야 하는 지침을 정의한다. 그 기술을 각각 살펴보면 다음과 같다.

3.1.1. 상호작용 UI 배치의 최적화

상호작용 UI 배치의 최적화는 안정적인 모션 입력을 위한 UI 배치 기준을 정의한다. 빛을 사용하는 프로젝트 매핑의 특성상 프로젝트 되는 광원을 마주 보고 [그림 3]과 같이 사용자가 있으면 사용자의 뒤쪽에는 그림자가 형성된다. 따라서 사용자에게 제시되는 UI 상호작용 버튼은 반드시 그림자가 생기지 않는 영역, 가급적 사용자의 전면에 위치해야 한다.



[그림 3] 상호작용 UI 배치를 통한 디스플레이 표현 최적화 기술 개념도



[그림 4] 사용자 행동에 의한 의사 확인 가이드라인 개념도

3.1.2. 사용자 입력을 위한 명령어 확인

사용자의 동작을 입력 수단으로 사용하는 '모션 컨트롤'의 특성 중 가장 빈번하게 발생하는 문제점은 '입력 오류'다. 이는 사용자의 의지와 무관하게 해당 영역의 입력 버튼이 동작을 하기에 나타나는 문제인데, 이 상황은 아무런 입력이 없었음에도 불구하고 동작인식 센서의 오동작으로 발생하거나, 혹은 사용자가 무심코 한 행동이 그 버튼 위로 지나쳤을 경우 발생된다.

사용자의 동작 입력이 없었을 때 발생하는 동작인식 센서의 오동작은 매우 드물게 발생하며, 동작인식 센서의 인식률을 높이면 자연스럽게 해결될 수 있다. 그러나 사용자가 무심코 한 행동으로 상호작용이 발생하는 것은 빈번하게 발생하는 오류로, 원인이 사용자에게 있기 때문에 이를 기술적으로 방지할 대책이 꼭 필요하다. 이러한 대책 중 하나가 [그림 4]와 같이 사용자가 해당 버튼을 명령어로 인식하고 확실하게 그 명령을 실시할 의지가 있음을 전달하는 것이다. 이러한 이유로 처음 표준이 제정될 2015년 당시에는 3초 이상 사용자가 해당 버튼을 누르고 있어야 한다고 정의하였다. 그러나, 동작 인식 센서의 정확도가 정교해지고 컴퓨팅 파워가 향상되면서 3초라는 시간은 너무 길고 불편한 요소가 됐다. 그래서 2020년 해당 표준을 개정하면서 3초를 0.5초로 조정했다.

3.1.3. 사용자 디스플레이 표현 최적화

'사용자 디스플레이 표현 최적화'에서 '사용자 디스플레이'는 사용자에게 제시되는 UI 인터페이스를 의미한다. 이 UI 인터페이스가 너무 크면 제한된 영역 안에서 콘텐츠가 제시해야 하는 다른 인터페이스와 이미지가 표현될 공간이 줄어들고, 반대로 너무 작으면 사용자의 동작 입력이 제대로 인식되지 못하는 오류가 발생한다. 이 역시 처음 표준이 제정된 2015년에는 사용자 신체 크기의 3배로 제시됐으나 2020년 개정판에서는 그간 향상된 정확도와 정교함을 반영하여 1.5배로 수정됐다.

3.2 깊이 카메라를 이용한 프로젝터의 인터페이스 매핑 지침

‘깊이 카메라를 이용한 프로젝터의 인터페이스 매핑 지침’ 표준은 2020년 7월 6일 실감형혼합 현실기술포럼의 포럼표준(IITF_01.0005)으로 제정된 이후, 12월 10일 TTA 단체 표준(TTAK.KO-10.1263)으로 제정됐다.

이 표준은 프로젝션 매핑으로 생성되는 인터페이스의 안정적인 작동을 보장하려는데 목적이 있다. 프로젝션 매핑 기술을 이용할 때 프로젝터는 프로젝션할 대상 영역의 범위에 따라 키스톤(keystone) 조절을 하는데, 이는 [그림 5]와 [그림 6]과 같이 프로젝션을 통해 제공되는 상호작용 인터페이스의 좌표값에 영향을 준다. 이 좌표값이 바뀌면 사용자의 동작이 인식되는 좌표값이 달라져야 하는데, 이 좌표값을 상대적으로 바꾸는 것은 현재 기술로 불가능하다. 따라서 상호작용 인터페이스의 좌표값이 키스톤에 의해서 변경되는 것을 방지해야만 한다.



[그림 5] 프로젝션 매핑에 사용되는 실제 크기의 이미지



[그림 6] 프로젝터의 키스톤 조절에 의하여 확대된 크기의 이미지

이와 관련하여 키스톤 조절로 상호작용 인터페이스의 좌표값 변화가 발생하지 않도록 상호작용 인터페이스가 있는 위치만 절대값으로 고정하여 프로젝션 이미지를 생성하는 기술을 주식회사 (주)조이편에서 개발했으며, 지적재산권이 출원(10-2019-0094261)된 상태에서 본 표준이 제정됐다.

3.3 사용자 정보 기반의 동작 출력 인터페이스 제작 참조 모델

‘사용자 정보 기반의 동작 출력 인터페이스 제작 참조 모델’ 표준 역시 ‘깊이 카메라를 이용한 프로젝터의 인터페이스 매핑 지침’ 표준과 같이 2020년 7월 6일 실감형혼합현실기술포럼의 포럼표준(IITF_01.0003)으로 제정된 이후 12월 10일 TTA 단체 표준(TTAK.KO-10.1251)으로 제정됐다. 이 표준은 사용자가 신체조건에 관계없이 편리하게 인터페이스를 이용하게 하는 데 목적이 있다.

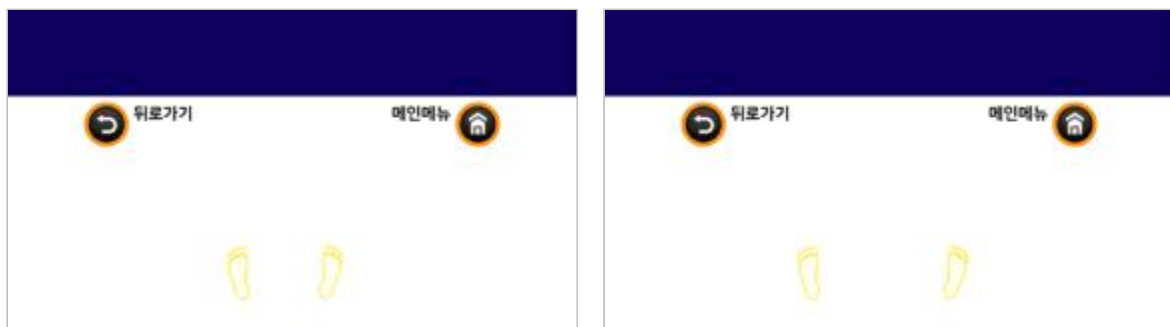
프로젝션 매핑 기술을 이용할 때 사용자와 상호작용이 용이하도록 인터페이스를 제공해야 하며, 사용자 편의성을 고려해 사용자가 불편 없이 이용할 수 있어야 한다. 예컨대 사용자의 보폭이 넓은 사람에게 지나치게 좁은 두 개의 인터페이스 간격을 제공하거나, 보폭이 좁은 사람에게 넓은 두 개의 인터페이스 간격을 제공한다면, 이는 사용자에게 매우 큰 불편을 줄 것이다.

이 표준은 ‘콘텐츠와 사용자 간 입출력 상호작용 지침’ 표준의 5.3항 ‘프로젝션 기반 콘텐츠 사용자 디스플레이 표현 최적화 기술’과 밀접한 관계가 있다. ‘프로젝션 기반 콘텐츠 사용자 디스플레이

레이 표현 최적화 기술'은 사용자의 신체와 관련하여 상호작용 인터페이스로 제공되는 버튼의 크기를 결정하는 것이라면, '사용자 정보 기반의 동작 출력 인터페이스 제작 참조 모델' 표준은 사용자의 신체 크기(size)에 따라 각 버튼 사이의 간격이 유기적으로 조정되는 것을 의미한다.

[그림 7]은 하프 스쿼트 동작을 가이드하는 콘텐츠에서 제공되는 인터페이스 예시다. 이 경우 신장 150cm와 180cm의 사용자에게 제시되는 보폭의 크기는 당연히 달라야 한다. 마찬가지로 [그림 8]의 푸쉬업 동작은 엮드리는 동작이므로 양팔과 양발 간격이 달라져야 한다.

이처럼 신체조건에 따라 인터페이스를 조절하는 것은 상식적으로 당연해 보이지만, 이와 관련된 표준이 없다면 실감형 혼합현실 콘텐츠 제작자들이 이러한 사항을 제작 과정에서 놓치거나 문제를 인식하더라도 해결 방향을 설정하는 데 어려움을 겪었을 것이다.



[그림 7] 신장 150cm(左)와 180cm(右) 사용자의 하프 스쿼트 보폭 변화



[그림 8] 신장 150cm(左)와 180cm(右) 사용자의 푸쉬업 양발과 양손 간격 변화

4. 맺음말

본고를 통해 실감형 혼합현실 콘텐츠 중 프로젝션 매핑 기술을 이용하는 콘텐츠의 제작에 필요한 몇 가지 기술 표준을 살펴보았다. 이 표준은 모두 TTA의 단체 표준으로 제정됐으며, 현재 실감형혼합현실기술포럼의 주도하에 IEEE 3079.2 TG에서 IEEE P3079.2 표준으로 제정이 진행 중이다.

최근 메타버스(metavers)가 주목받으면서 XR(eXtended Reality)라고 하는 개념도 많은 이의 관심을 끌고 있다. XR이 신개념처럼 보이지만 실은 혼합현실(MR)과 가상현실(VR)을 통합한 것에 불과하다. 정확히 말하면 개념적으로 혼합현실과 가상현실을 합치는 것은 각각의 기본 정의를 부정하는 것이기 때문에 불가능하고, 그저 혼합현실과 가상현실 두 개념을 통칭한다는 것이 명확한 표현일 것이다. 본고에서는 XR의 한 축을 담당하고 있는 MR, 그중에서도 실감형

혼합현실을 표현하는 콘텐츠 기술에 가장 기초적이고 핵심적인 기술을 다뤘다.

요즘 많은 곳에 [그림 9]처럼 프로젝션 매핑을 이용한 아쿠아리움이 어린이를 위해 설치되어 있다. 이 콘텐츠의 묘미는 아이들의 움직임과 상호작용(interaction)하는 물고기의 모습이다. 이러한 상호작용이 원활하게 이루어지려면 앞서 살펴본 기술들이 잘 반영되어야 한다. 본고에서 살펴본 표준이 이처럼 우리 생활 속 가까운 곳에 폭넓게 활용되는 만큼, 실감형 혼합현실 기술에 대한 지속적인 관심이 필요하다.



[그림 9] 프로젝션 매핑을 이용한 아쿠아리움[4]

[주요 용어 풀이]

- 혼합현실(MR, Mixed Reality): 가상의 객체(virtual object)와 실제의 객체(physical object)가 자연스럽게 상호작용할 수 있도록 하는 기술로 구현되는 콘텐츠 기술
- 프로젝션 매핑(Projection mapping): 실제의 객체에 프로젝션을 통해 만들어진 가상의 객체 이미지를 투영함으로써 혼합현실을 구성하는 기법

[참고문헌]

- [1] '체감형 모션게임 Xtion, 국내 출시 기념 가격인하 이벤트',
(<http://biz.heraldcorp.com/view.php?ud=20131216000746>), 2013.12.16.
- [2] '가상현실과 혼합현실을 정확히 알아야 하는 이유', 정상권, 파이낸셜신문,
(<http://www.efnews.co.kr/news/articleView.html?idxno=81594>), 2019.10.30.
- [3] Kinect v1, v2 그리고 Azure Kinect 비교, DEVELOPER, (<https://ctkim.tistory.com/78>), 2020.07.25.
- [4] 프로젝션 매핑_아쿠아리움, (<https://www.youtube.com/watch?v=ezxC7dBh0cw>), 2021.04.06.

※ 출처: TTA 저널 제195호

(코로나 이슈로 각 표준화기구의 표준화회의가 연기·취소됨에 따라 TTA 저널로 대체합니다)