

개발 동기 및 목적

Vision Pro의 편리하고 직관적인 제스처에 신선한 충격을 받았다. 그러나 우리가 흔히 사용하는 웹과 앱 환경에서는 이러한 직관적인 제스처 기술이 크게 활성화되어 있지 않는데, 그 이유는 일반 카메라의 한계와 웹에 쉽게 적용할 수 있는 모듈이나 라이브러리가 없기 때문이라고 느꼈다. 따라 이번 프로젝트에서는 일반 카메라 프레젠테이션을 원거리에서 제스처로 진행할 수 있는 Desktop 앱 'Gesto'를 개발하였다. 제스처 인터페이스의 가능성과 한계점을 알아보고, 나아가 이를 웹 기반에서 쉽게 사용할 수 있도록 모듈화하여 개발하는 것을 목표로 한다.

주요 기술

1 Electron.js

- 크로스플랫폼 지원
- 동일한 소스로 웹서비스로 배포 가능

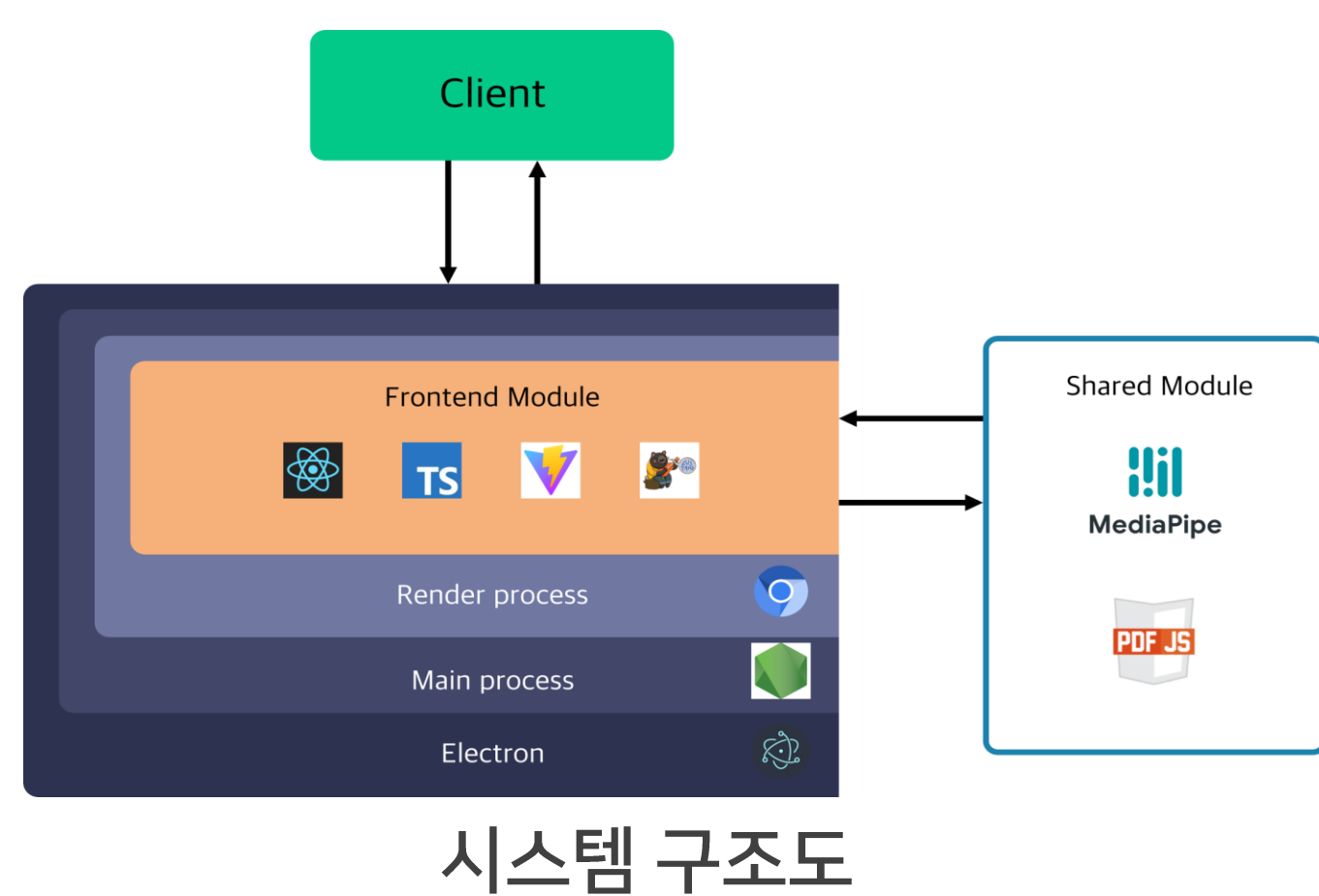
2 Vite

- HMR을 통한 개발 환경 개선
- ES module을 통한 cold start 시간 단축

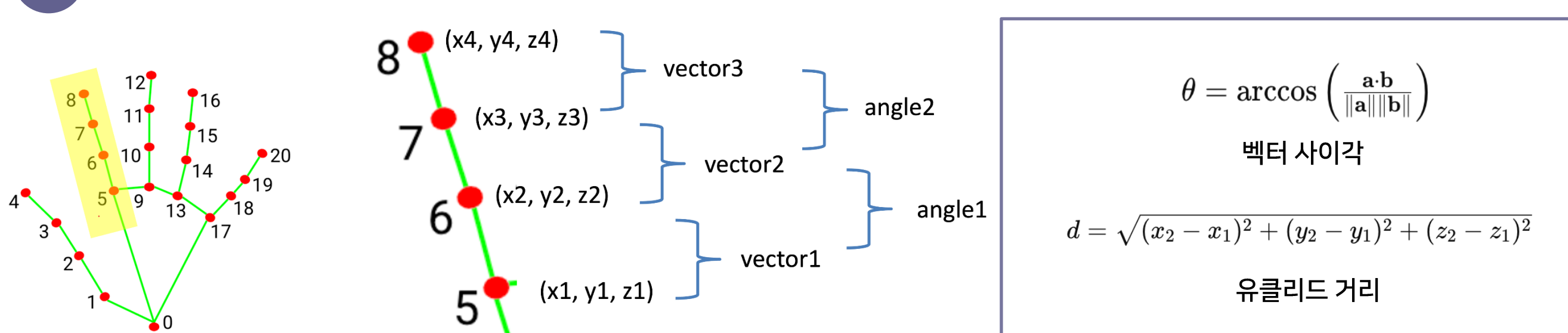
3 MediaPipe Hand Landmarks Solution

- 손을 감지하고 3차원 좌표상에 랜드마크를 예측
- 최적화된 성능으로 인해 프레임 단위 호출에 적합

개발 내용

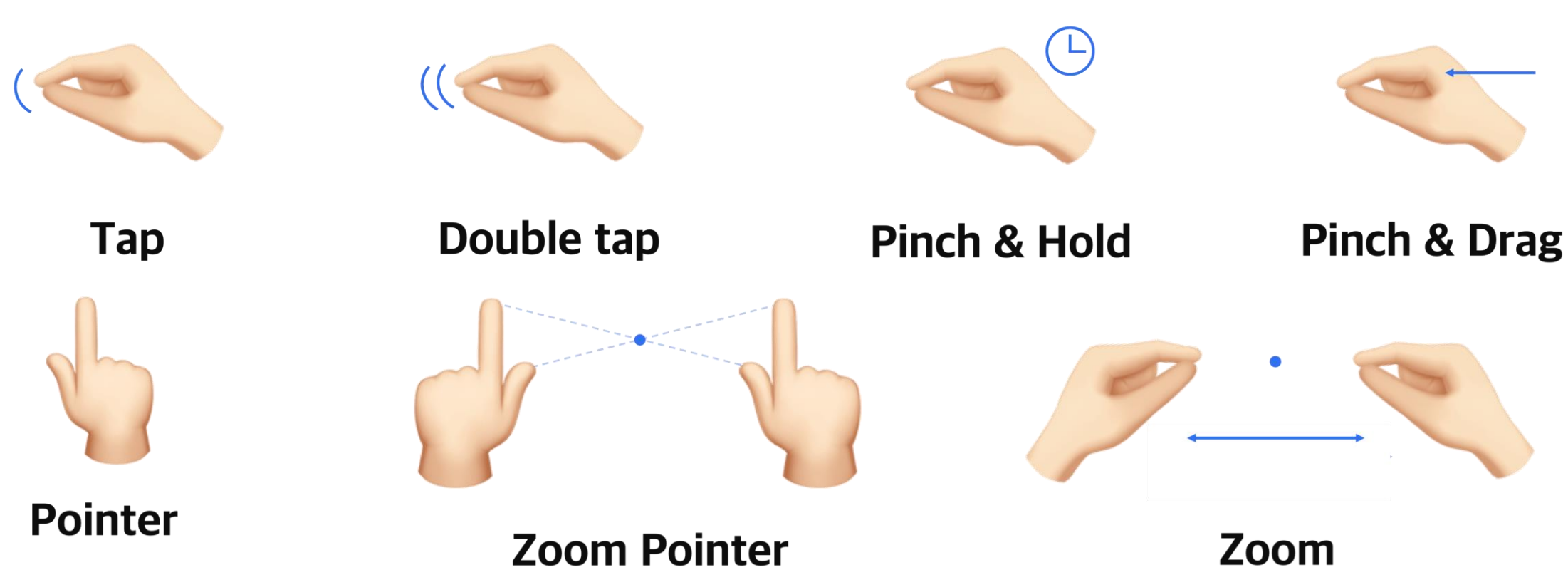


1 제스처 인식



- ◆ Hand Landmarks Solution를 통해 21개의 3차원 랜드마크를 감지, 벡터 사이각을 구하여 손가락이 꺾인 각도 계산
- ◆ 유클리드 거리 공식을 사용하여 3차원 상에서 두 점 사이 거리 계산
- ◆ 각도, 손가락 번호, 거리값을 바탕으로 손가락의 곧게 펴짐, 접힘, 접촉 여부 등을 검사할 수 있는 로직 모듈화

구현한 제스처는 아래와 같다.



2 인식률 개선

일반 카메라, 빠른 움직임에서의 제스처 인식률을 높이기 위해 다음과 같은 방법을 사용하였다.

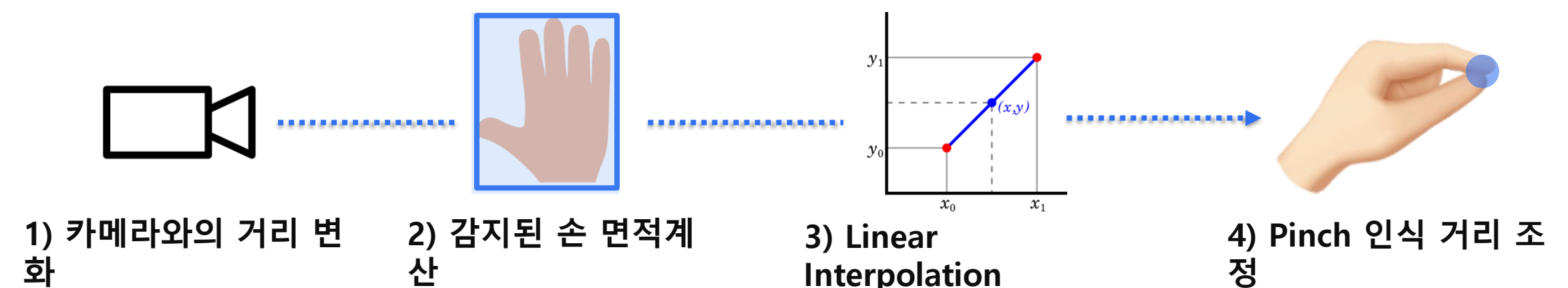
1) History, 속도 기반 제스처 변경 제어

- ◆ 제스처를 결정할 때, 여러 번 동일한 값을 감지했을 때 변경한다.
- ◆ 현재 프레임에서의 좌표와 이전 프레임에서의 좌표를 기반으로 실시간으로 손이 움직이고 있는 속도를 추적하고, 이를 기반으로 제스처 변경을 제어한다.

이를 통해 일시적 오류나 노이즈로 인한 제스처 변경을 방지하고 움직이는 상태를 바탕으로 하는 Pinch & Drag(잡고 끌기), Zoom(확대 및 축소) 제스처 인식을 개선했다.

2) Linear Interpolation을 통한 인식 거리 조정

랜드마크(점) 기반이기 때문에, 손가락 접촉을 판단할 때 카메라와의 거리에 따라 오차가 발생한다. 손 면적과 인식 거리를 기반으로 Linear Interpolation을 수행해 인식 거리를 실시간으로 조정해 거리 변화와 관계없이 동일한 성능을 보일 수 있도록 구현했다



3) 제스처의 웹 이벤트 변환 및 핸들링

1) 판단한 제스처를 웹 이벤트와 연결

Window와 동일한 사이즈의 투명한 canvas를 최상단 레이어에 배치하고 커서를 렌더링했다. canvas의 pointerEvents 속성의 값을 none으로 고정시켜 event가 레이어 아래 element에 전달된다. 제스처를 웹 이벤트로 변환하고, 좌표와 elementFromPoint 메서드를 통해 target element가 결정되면 변환된 웹이벤트가 dispatch된다.

2) Click, Drag

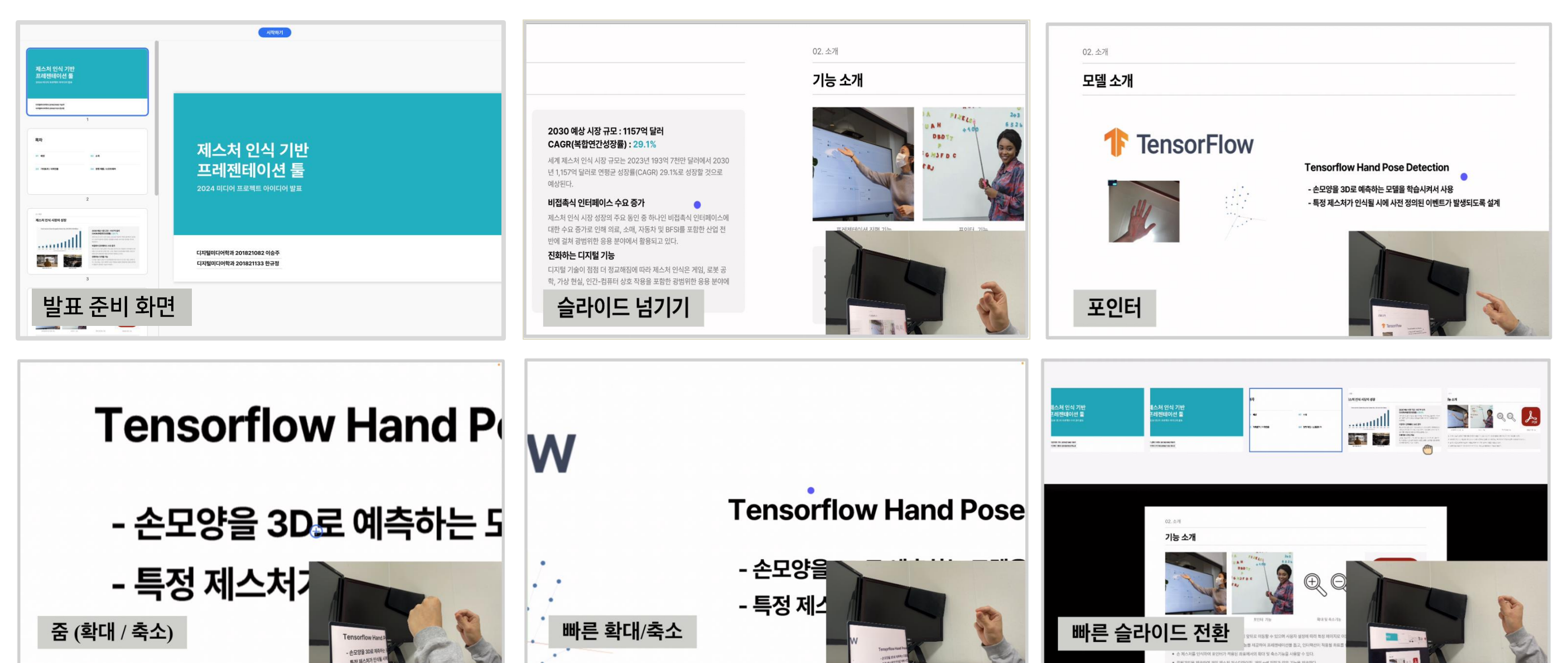
마우스에서는 Click, Drag를 mouse down과 up이 동일한 위치에서 발생했는지에 따라 구분한다. 그러나 손은 미세하게 계속 움직이기에, Click에 대한 기준을 다르게 적용했다. 마지막 제스처와 현재 제스처를 비교해 Pinch 유지 시간을 체크하고, 기준 시간을 넘기기 전에는 move 이벤트를 발생시키지 않도록 하여 Click과 Drag를 구분했다.

3) Zoom

유클리드 거리로 구한 양손 좌표의 사이 거리에 따라 target element에서 양손의 중앙 좌표를 transform origin으로 하는 확대/축소가 일어나도록 구현했다.

결과 및 분석

손으로만 조작할 수 있는 제한된 환경에서, 사용성을 최대화 할 수 있는 터치스크린과 비슷한 방식의 제스처 인터페이스를 구현하였고, 실제로 원거리에서 제어 가능한 프레젠테이션 툴을 개발하였다. 구현한 주요 기능은 다음과 같다.



1800프레임 기준 인식 테스트를 진행한 결과 모든 제스처에서 100%에 가까운 인식률을 보였으며, 빠르게 움직이는 환경에서도 높은 인식률을 보였다.

결과적으로, 일반 카메라를 사용하는 웹/앱에서 제스처 인터페이스의 가능성을 확인할 수 있었다. 한계점으로는 빛이 아예 없는 환경에서는 사용하기 힘들고, 손이 미세하게 흔들리기에 정밀한 작업에는 적용이 어렵다는 점이 있었다.

향후 계획

현재 구현된 프로토타입을 다양한 환경에서의 테스트를 진행하고 리팩토링 후 앱과 웹서비스로 출시할 예정이다. 또한 제스처 인식 모듈을 웹 기반에서 활용할 수 있도록 패키징하여 오픈소스로 배포할 예정이다.



오픈소스 URL / 시연영상

<https://github.com/gesto-project/GestoApp>

