


Face mesh



MediaPipe Face Mesh

Overview

MediaPipe Face Mesh는 모바일 장치에서도 실시간으로 468개의 3D 얼굴 랜드마크를 추정하는 솔루션입니다. 기계 학습(ML)을 사용하여 3D 얼굴 표면을 추론하며, 전용 깊이 센서 없이 단일 카메라 입력만 필요합니다. 파이프라인 전체의 GPU 가속과 함께 가벼운 모델 아키텍처를 활용하는 이 솔루션은 실시간 경험에 중요한 실시간 성능을 제공합니다.

또한, 이 솔루션은 얼굴 랜드마크 추정과 유용한 실시간 증강 현실(AR) 애플리케이션 사이의 격차를 해소하는 얼굴 변환 모듈과 번들로 제공됩니다. 메트릭 3D 공간을 설정하고 얼굴 랜드마크 화면 위치를 사용하여 해당 공간 내에서 얼굴 변환을 추정합니다. 얼굴 변환 데이터는 얼굴 포즈 변환 매트릭스와 삼각형 얼굴 메쉬를 포함한 일반적인 3D 프리미티브로 구성되어 있다. 후드 아래에서, Procrustes Analysis라는 가벼운 통계 분석 방법이 강력하고 성능이 뛰어나며 휴대용 논리를 구동하기 위해 사용된다. 분석은 CPU에서 실행되며 ML 모델 추론 위에 최소한의 속도/메모리 공간을 가지고 있다.

ML Pipeline

우리의 ML 파이프라인은 함께 작동하는 두 가지 실시간 심층 신경망 모델로 구성되어 있습니다: 전체 이미지에서 작동하고 얼굴 위치를 계산하는 검출기와 해당 위치에서 작동하고 회귀를 통해 대략적인 3D 표면을 예측하는 3D 얼굴 랜드마크 모델. 얼굴을 정확하게 자르면 회전, 번역 및 스케일 변경으로 구성된 아핀 변환과 같은 일반적인 데이터 증강의 필요성을 크게 줄일 수 있습니다. 대신 네트워크가 예측 정확도를 조정하기 위해 대부분의 역량을 바칠 수 있게 해준다. 또한, 파이프라인에서 작물은 이전 프레임에서 식별된 얼굴 랜드마크를 기반으로 생성될 수 있으며, 랜드마크 모델이 더 이상 얼굴 존재를 식별할 수 없는 경우에만 얼굴을 다시 localize하기 위해 호출된 얼굴 탐지기가 호출됩니다. 이 전략은 손 랜드마크 모델과 함께 손바닥 탐지기를 사용하는 MediaPipe Hands 솔루션에 사용된 것과 유사합니다.

파이프라인은 얼굴 랜드마크 모듈의 얼굴 랜드마크 서브그래프를 사용하고 전용 페이스 렌더러 서브그래프를 사용하여 렌더링하는 MediaPipe 그래프로 구현됩니다. 얼굴 랜드마크 서브그래프는 내부적으로 얼굴 감지 모듈에서 `face_detection_subgraph`를 사용합니다.

참고: 그래프를 시각화하려면 그래프를 복사하여 MediaPipe Visualizer에 붙여넣으세요. 관련 하위 그래프를 시각화하는 방법에 대한 자세한 내용은 시각화 문서를 참조하십시오.

Models

FACE DETECTION MODEL

얼굴 탐지기는 MediaPipe 얼굴 감지에 사용된 것과 동일한 BlazeFace 모델이다. 자세한 내용은 MediaPipe 얼굴 감지를 참조하십시오.

FACE LANDMARK MODEL

3D 얼굴 랜드마크의 경우 전송 학습을 사용하고 여러 목표를 가진 네트워크를 훈련시켰습니다: 네트워크는 합성 렌더링 데이터의 3D 랜드마크 좌표 좌표와 주석이 달린 실제 데이터에 대한 2D 의미론적 윤곽을 동시에 예측합니다. 그 결과 네트워크는 우리에게 합성뿐만 아니라 실제 데이터에 대한 합리적인 3D 획기적인 예측을 제공했다.

3D 랜드마크 네트워크는 추가 깊이 입력 없이 자른 비디오 프레임을 입력으로 받습니다. 이 모델은 3D 포인트의 위치와 얼굴이 존재하고 입력에 합리적으로 정렬될 확률을 출력한다. 일반적인 대안적인 대안적인 접근 방식은 각 랜드마크에 대한 2D 히트맵을 예측하는 것이지만, 깊이 예측에 적합하지 않으며 많은 점에 대한 계산 비용이 높다. 우리는 예측을 반복적으로 부트스트랩하고 정제함으로써 모델의 정확성과 견고성을 더욱 향상시킵니다. 그렇게 하면 데이터 세트를 핑그림, 비스듬는 각도 및 폐색과 같은 점점 더 도전적인 사례로 성장시킬 수 있습니다.

이 논문에서 얼굴 랜드마크 모델에 대한 자세한 정보를 찾을 수 있습니다.

ATTENTION MESH MODEL

얼굴 랜드마크 모델 외에도 우리는 의미론적으로 의미 있는 얼굴 영역에 주의를 기울이는 또 다른 모델을 제공하므로 더 많은 계산을 희생시키면서 입술, 눈 및 홍채 주변의 랜드마크를 더 정확하게 예측합니다. 그것은 AR 메이크업과 AR 인형극과 같은 애플리케이션을 가능하게 한다.

주의 메쉬 모델은 refine_landmarks 옵션을 통해 솔루션 API에서 선택할 수 있습니다. 이 논문에서 모델에 대한 더 많은 정보를 찾을 수 있습니다.

Face Transform Module

얼굴 랜드마크 모델은 화면 좌표 공간에서 단일 카메라 얼굴 랜드마크 감지를 수행합니다: X 및 Y 좌표는 정규화된 화면 좌표이며, Z 좌표는 상대적이며 약한 원근법 투영 카메라 모델에서 Xcoordinate로 확장됩니다. 이 형식은 일부 응용 프로그램에 적합하지만, 가상 3D 물체를 감지된 얼굴로 정렬하는 것과 같은 증강 현실(AR) 기능의 전체 스펙트럼을 직접 활성화하지는 않습니다.

얼굴 변환 모듈은 화면 좌표 공간에서 미터법 3D 공간으로 이동하며 감지된 얼굴을 일반 3D 물체로 처리하는 데 필요한 프리미티브를 제공합니다. 설계상, 원근법 카메라를 사용하여 얼굴 랜드마크 위치가 변경되지 않도록 최종 3D 장면을 화면 좌표 공간에 다시 투사할 수 있습니다.

Key Concepts

미터법 3D 공간

얼굴 변환 모듈 내에 설정된 메트릭 3D 공간은 오른손잡이 직교 메트릭 3D 좌표 공간이다. 공간 안에는 우주 원점에 위치한 가상 원근법 카메라가 있으며 Z축의 음수 방향을 가리킨다. 현재 파이프라인에서 입력 카메라 프레임은 정확히 이 가상 카메라에 의해 관찰되므로 매개 변수는 나중에 화면 랜드마크 좌표를 미터법 3D 공간으로 다시 변환하는 데 사용됩니다. 가상 카메라 매개 변수는 자유롭게 설정할 수 있지만, 더 나은 결과를 얻으려면 가능한 한 실제 물리적 카메라 매개 변수에 가깝게 설정하는 것이 좋습니다.

표준 얼굴 모델

캐노니컬 페이스 모델은 얼굴 랜드마크 모델의 468 3D 얼굴 랜드마크 토폴로지를 따르는 인간 얼굴의 정적 3D 모델이다. 이 모델은 두 가지 중요한 기능을 갖추고 있다:

메트릭 단위 정의: 표준 면 모델의 스케일은 메트릭 3D 공간의 메트릭 단위를 정의합니다. 기본 표준 면 모델이 사용하는 미터법 단위는 센티미터이다.

브리지 정적 및 런타임 공간: 얼굴 포즈 변환 매트릭스는 실제로 표준 얼굴 모델에서 각 프레임에 추정된 런타임 페이스 랜드마크 세트로의 선형 맵입니다. 이렇게 하면, 표준 얼굴 모델을 중심으로 모델링된 가상 3D 자산은 얼굴 포즈 변환 매트릭스를 적용하여 추적된 얼굴과 정렬할 수 있습니다.

Components

파이프라인을 변형시키십시오

변환 파이프라인은 미터법 3D 공간 내의 얼굴 변환 물체를 추정하는 핵심 구성 요소이다. 각 프레임에서, 다음 단계는 주어진 순서로 실행됩니다:

얼굴 랜드마크 화면 좌표는 미터법 3D 공간 좌표로 변환됩니다.

얼굴 포즈 변환 매트릭스는 표준 얼굴 메트릭 랜드마크에서 둘 간의 차이를 최소화하는 방식으로 설정된 런타임 얼굴 메트릭 랜드마크로 엄격한 선형 매핑으로 추정됩니다.

얼굴 메시는 런타임 얼굴 메트릭 랜드마크를 정점 위치(XYZ)로 사용하여 생성되는 반면, 정점 텍스처 좌표(UV)와 삼각형 토폴로지는 모두 표준 얼굴 모델에서 상속됩니다.

변환 파이프라인은 MediaPipe 계산기로 구현됩니다. 귀하의 편의를 위해, 이 계산기는 해당 메타데이터와 함께 통합 MediaPipe 하위 그래프에 번들로 제공됩니다. 얼굴 변환 형식은 프로토콜 버퍼 메시지로 정의됩니다.

효과 렌더링

효과 렌더러는 얼굴 효과 렌더러의 작업 예 역할을 하는 구성 요소이다. 모바일 장치에서 실시간 성능을 활성화하기 위해 OpenGL ES 2.0 API를 대상으로 하며 다음과 같은 렌더링 모드를 지원합니다:

3D 객체 렌더링 모드: 가상 물체는 얼굴에 부착된 물체를 에뮬레이트하기 위해 감지된 얼굴과 정렬됩니다(예: 안경);

페이스 메시 렌더링 모드: 페이스 페인팅 기술을 모방하기 위해 페이스 메시 표면 위에 텍스처가 늘어났다.

두 렌더링 모드에서, 페이스 메시는 먼저 깊이 버퍼로 바로 오클루저로 렌더링됩니다. 이 단계는 얼굴 표면 뒤에 보이지 않는 요소를 숨겨서 더 믿을 수 있는 효과를 만드는 데 도움이 된다.

효과 렌더러는 MediaPipe 계산기로 구현됩니다.

Solution APIs

Configuration Options

명명 스타일과 가용성은 플랫폼/언어마다 약간 다를 수 있습니다.

STATIC_IMAGE_MODE

False로 설정하면, 솔루션은 입력 이미지를 비디오 스트림으로 취급합니다. 그것은 첫 번째 입력 이미지에서 얼굴을 감지하려고 시도할 것이며, 성공적으로 감지되면 얼굴 랜드마크를 더 현지화합니다. 후속 이미지에서, 모든 max_num_faces 얼굴이 감지되고 해당 얼굴 랜드마크가 현지화되면, 얼굴의 추적을 잃을 때까지 다른 탐지를 호출하지 않고 랜드마크를 추적하기만 하면 됩니다. 이것은 대기 시간을 줄이고 비디오 프레임 처리에 이상적입니다. True로 설정하면 얼굴 감지는 모든 입력 이미지에서 실행되며, 관련이 없을 수 있는 정적 이미지 배치를 처리하는 데 이상적입니다. 기본값은 false입니다.

MAX_NUM_FACES

감지할 최대 얼굴 수. 기본값은 1입니다.

REFINE_LANDMARKS

눈과 입술 주위의 랜드마크 좌표를 더욱 구체화하고, 주의 메쉬 모델을 적용하여 홍채 주변에 추가 랜드마크를 출력할지 여부. 기본값은 false입니다.

MIN_DETECTION_CONFIDENCE

감지 성공으로 간주될 얼굴 감지 모델의 최소 신뢰 값([0.0, 1.0]). 기본값은 0.5입니다.

MIN_TRACKING_CONFIDENCE

얼굴 랜드마크가 성공적으로 추적되는 것으로 간주될 랜드마크 추적 모델의 최소 신뢰 값([0.0, 1.0]) 또는 그렇지 않으면 얼굴 감지가 다음 입력 이미지에서 자동으로 호출됩니다. 더 높은 값으로 설정하면 더 높은 대기 시간을 희생시키면서 솔루션의 견고성을 높일 수 있습니다. Static_image_mode가 참이면 무시되며, 얼굴 감지는 단순히 모든 이미지에서 실행됩니다. 기본값은 0.5입니다.

Output

명명 스타일은 플랫폼/언어마다 약간 다를 수 있습니다.

MULTI_FACE_LANDMARKS

감지/추적된 얼굴 컬렉션, 각 면은 468개의 얼굴 랜드마크 목록으로 표시되며 각 랜드마크는 x, y 및 z로 구성됩니다. x와 y는 각각 이미지 너비와 높이에 의해 [0.0, 1.0]으로 정규화됩니다. z는 머리 중앙의 깊이가 원점인 랜드마크 깊이를 나타내며, 값이 작을수록 랜드마크가 카메라에 더 가깝습니다. Z의 크기는 x와 거의 같은 스케일을 사용한다.