Real-Time Facial Motion Capture & Animation on Mobile

Emiliano Gambaretto, Director of Research
Stefano Corazza, CEO
Character Animation and the Power of Storytelling
Over 100m people go watch animated movies/year
Animating Characters in 3D is hard

High-end Keyframe Animation
Animating Characters in 3D is hard

Using Motion Capture
What if it was as easy as doing a video chat?
Video chat using a 3D character? How fun would that be?
Face Plus - Technology Breakthrough
Face Plus: Movie Quality

- Award Winning Animated Short “Unplugged”
- Real-Time, Works with a standard webcam
Face Plus: Movie Quality

- Award Winning Animated Short “Unplugged”
- Real-Time, Works with a standard webcam
Machine Learning is the Enabler

Productivity and ease of use

Pure Manual (e.g. Disney old days)

Digital Tools (Deterministic)

Data-driven Interactive Content (Machine Learning)

Level of specialization

1900s

2000s

2015
Computer Vision vs Computer Graphics

• **Computer Graphics**

  - Expression & Emotions
  - Geometry
  - Texture

• **Computer Vision**

  - Expression & Emotions
  - Texture
  - Geometry
Machine learning applied to facial animation

Learning a map between a face appearance and the expression/emotion it conveys
At runtime the trained algorithm:

- Detect faces
- Extract facial expressions and emotions
- Provide the rendering engine with the information required to animate a character
Faceplus processing pipeline

Face Detection
Viola Jones framework\textsuperscript{1,2}

Face Tracking

Expression Mapping

Retargeting/animation

\textsuperscript{1}Robust real-time object detection. Paul Viola, Michael Jones. IJCV 2001

\textsuperscript{2}Find a Face. Alexander Lyashevsky. AMD Fusion developer summit 2011
Faceplus processing pipeline

Face Detection
Viola Jones framework\textsuperscript{1,2}

Face Tracking

Expression Mapping

Retargeting/animation

\textsuperscript{1}Robust real-time object detection. Paul Viola, Michael Jones. IJCV 2001
\textsuperscript{2}Find a Face. Alexander Lyashevsky. AMD Fusion developer summit 2011
Faceplus: face tracking

**Global Models**

Deformable model that aims at modeling the appearance of the entire face.

The model parameters are determined by matching the model to the input image.

**Pros**
- Accurate
- Well suited for GPU acceleration

**Cons**
- Inverse problem subject to local minima
- Require initialization / re-initializations
- Computationally intensive

*Active Appearance Models. T.F. Cootes, G.J. Edwards, C.J. Taylor. ECCV 1998*
Faceplus: face tracking

Global Models

Deformable model that aims at modeling the appearance of the entire face
The model parameters are determined by matching the model to the input image

Pros
- Accurate
- Well suited for GPU acceleration

Cons
- Inverse problem subject to local minima
- Require initialization / re-initializations
- Computationally intensive

Texture Component

\[
T = A [c, 1]
\]
Face global modeling

\[ M = W(V, V_{uv}, T) \]

Putting the global model together
Faceplus: face tracking

Local Models

Face modeled as a set of rigid 2D patches
Local features are detected separately
Local features locations are used to reconstruct the facial expression

Pros
- Robust
- Doesn’t require any initialization
- Highly parallelizable / Scalable

Cons
- Noisy reconstruction
- Computationally intensive
- Tricky SIMD implementation

Texture Component

Geometric Component

- Each patch is associated with a distribution model that determines its location with respect to the detected face
How to invert these models?
Inverting the face global model

Obtaining a residual to be minimized in a least square sense

\[ r = W \left( V_{uv}, V, I_i \right) - T \]
Inverting the face global model

Inverting the model

- Generate the geometric model (vertices locations) and the texture model (image) from the model parameters
- Each working item processes one pixel in the texture space:
  - Computes the pixel location in the video input space
  - Samples the video input at the computed location → input video has to be an image buffer
  - Subtracts the corresponding texture model
  - Residual norm computed through reduction
  - Images in the texture space don't need to undergo sampling operations → Memory Buffers
Inverting the face global model in an heterogeneous system

• The registration is achieved by minimizing the residual in a least square sense.

• The GPU takes care of the most computationally intensive and highly parallel tasks (computation of the residual, Jacobian and some linear algebra)

• The CPU takes care of tasks that are harder to parallelize. (solution of small, constrained linear systems)

• AAM inverse problem is notably slow on CPU. Most methods presented in literature to speed it up involve approximations that affects the robustness and efficiency of the method.

• Alternatively the AAM inverse problem can be solved very efficiently on heterogeneous systems.
Inverting the face local model in an heterogeneous system

- **(CPU)** The geometric model is used to determine the possible location of patches given the location of the detected face.
- **(GPU)** Features detectors are used to located the patches in the input image.
- **(CPU)** A clusterization algorithm takes care of grouping multiple detections.
- **(CPU)** A robust reconstruction algorithm estimates the geometric model (yellow line) given the detected patches (yellow dots).
- **(GPU)** The global model is used to refine the model estimation (blue line).
Implementation

• Faceplus is a library.
• Faceplus is portable: it is written in C++ and OpenCL.
• Faceplus scales with the hardware.
• Faceplus can be used within 3rd party software: Unity3d, Motion Builder.
Conclusions

- Animation Creation reaching consumer
- Very computationally intensive
- Modern platforms such as the Tegra-K1 enable it on mobile
- Potential for real-time applications such as video chat 2.0
Thank You!