# Motion capturing via computer vision and machine learning

A live demonstration & talk

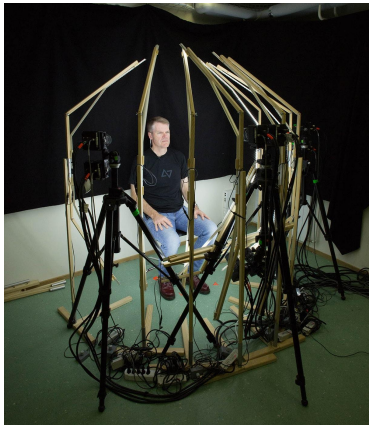
Justin Tennant
June 13, 2017
San Jose State University

# What is Stringless?

- Facial motion capture (“expression tracking”) system
- Integrates directly into Autodesk Maya
- Extendable to other animation software
- Undemanding system requirements, minimal hardware requirements
  - Intended user: casual/non-professional animator, visdev purposes
- A prototype, not a complete product

# Motion capture at a glance

- Digital recording of person's facial movements, applied to CG character's face
- Markered vs markerless tracking
- 2D vs 3D (single camera vs multi-camera, respectively)
- Realtime vs capture-only

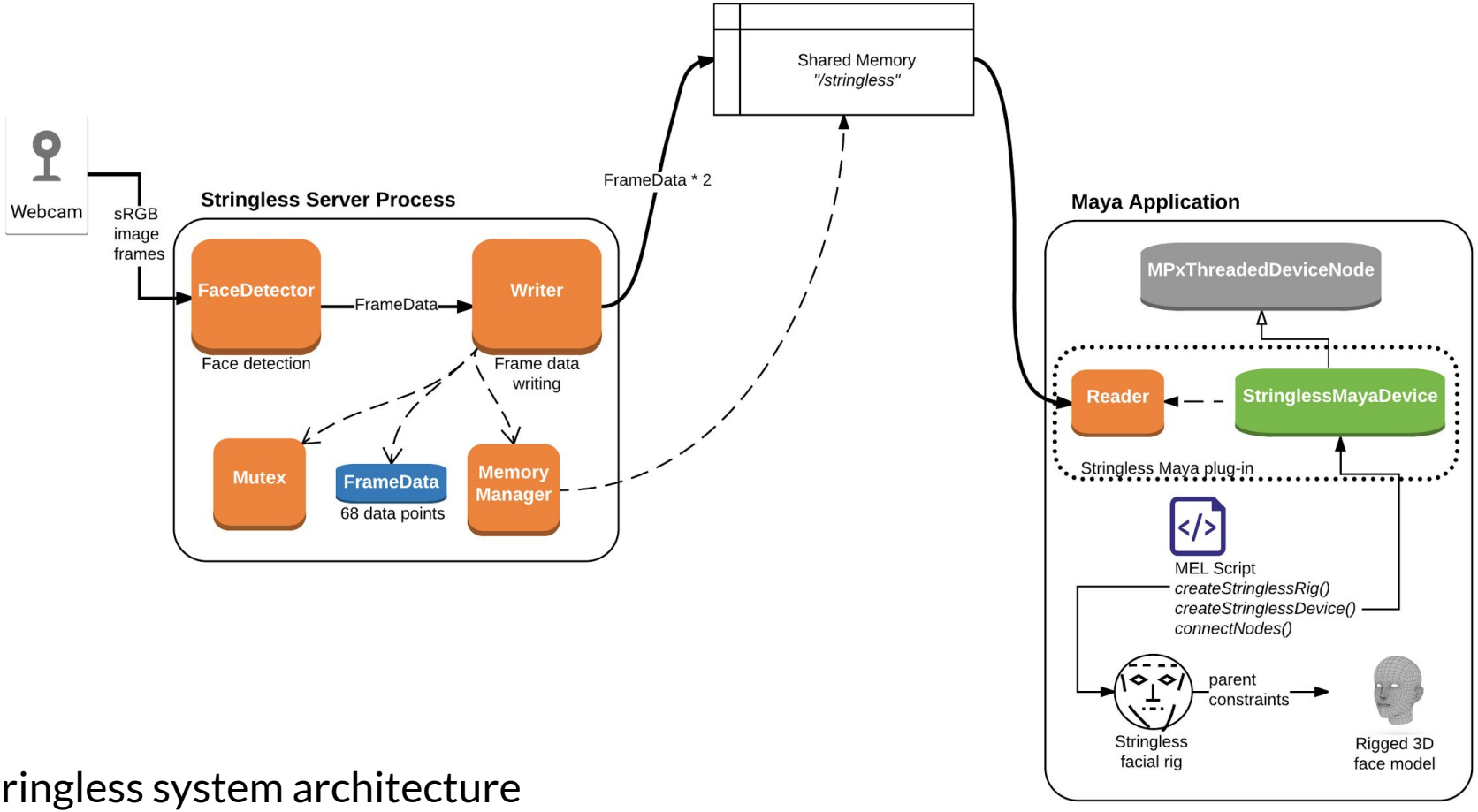
Facial motion capture systems, from top left to bottom left, clockwise:
*Unnamed Avatar Capture System (Weta Digital), Facerig (Holotech Studios), Faceshift (faceshift AG), Medusa (Disney Research)*

# Defining characteristics of Stringless

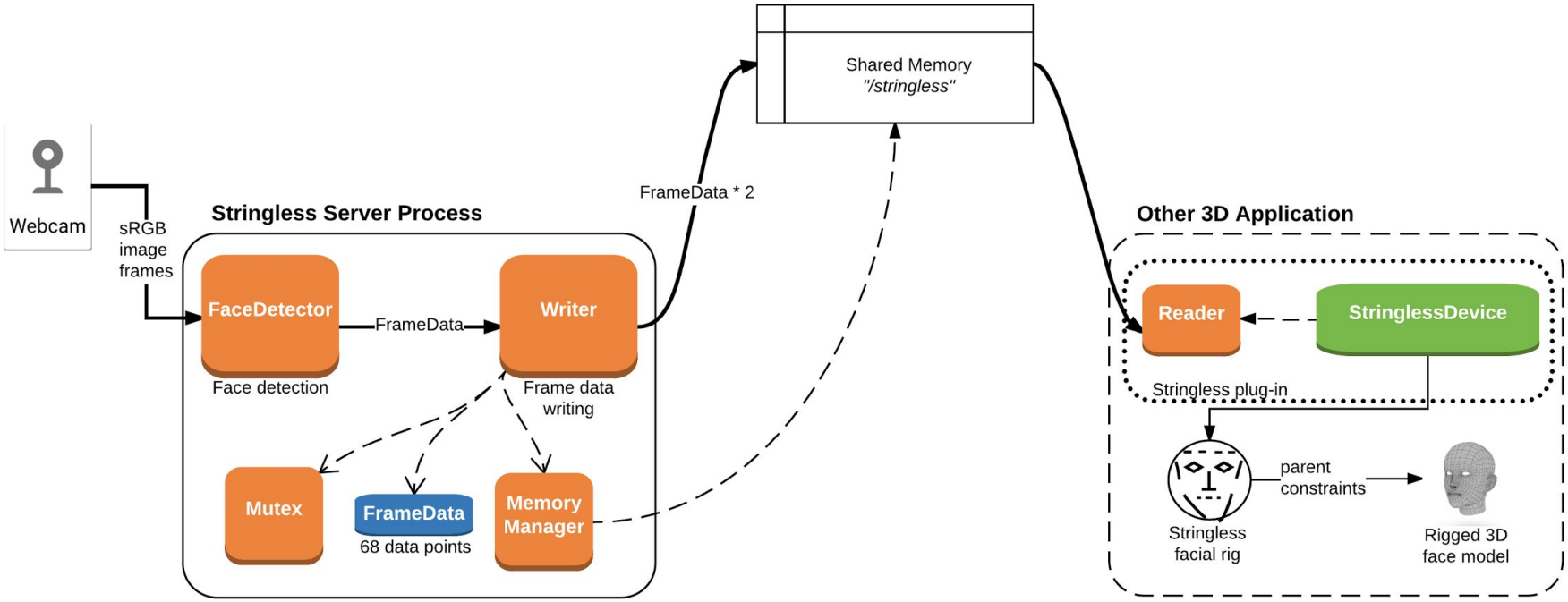
- Markerless
- Monocular camera input
- Real-time

# Stringless operation flow

1. Server process:
    a. Captures image frames from webcam
    b. Detects a face in frame
        i. Locate facial landmark positions
    c. Sends landmark data (*FrameData*) to shared memory
2. Maya plug-in:
    a. Creates a unique face rig separate from the 3D rigged face model
    b. Creates a device to read live data from shared memory
    c. Links device data to face rig joints
    d. User selectively constrains 3D face model controls to face rig




# Stringless system architecture

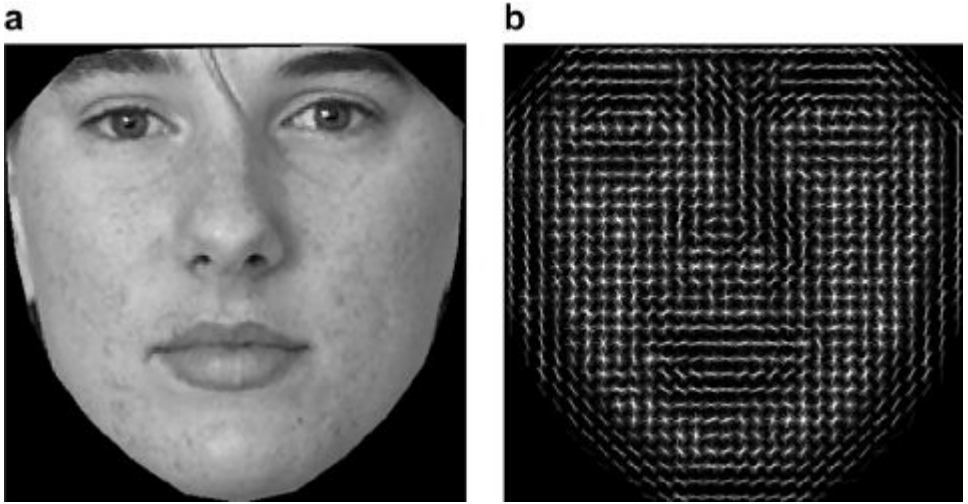



# Stringless system architecture

# Face Detection Technology

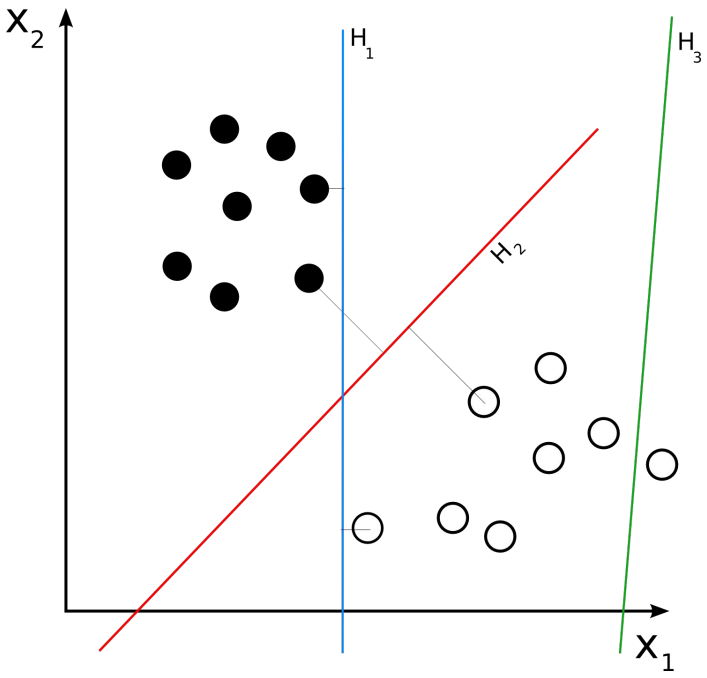
Implemented with dlib machine learning toolkit

- Histogram of Oriented Gradients (HOG)
- Linear classifier
- Image pyramid
- Sliding window detection scheme




https://www.researchgate.net/figure/284243500_fig5_Fig-5-Histogram-of-Oriented-Gradients-HOG-features-in-4-4-cells

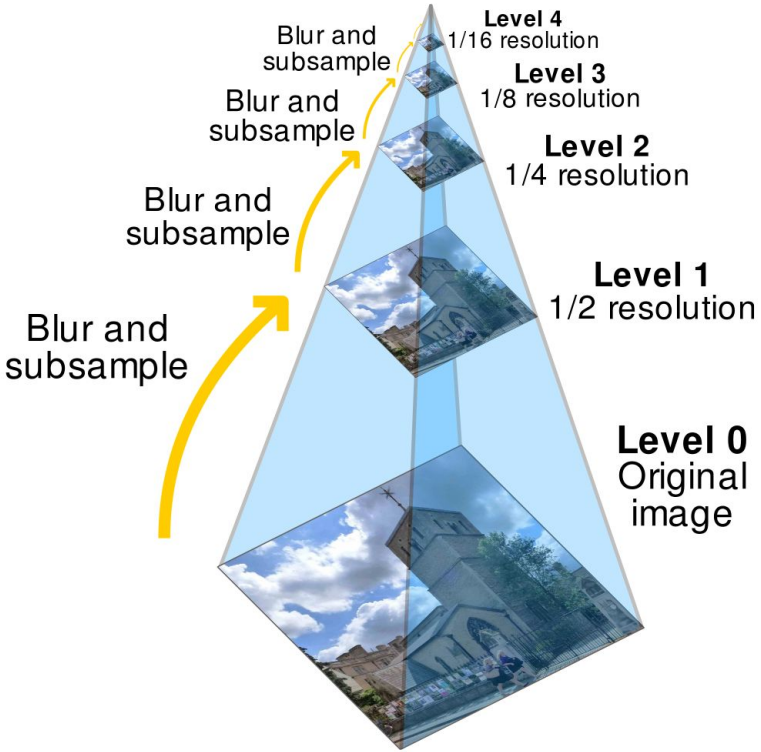
# Histogram of Oriented Gradients




iBUG 300 Faces In-The-Wild Challenge (300-W), IMAVIS 2014

C. Sagonas, E. Antonakos, G, Tzimiropoulos, S. Zafeiriou, M. Pantic. **300 faces In-the-wild challenge: Database and results**. Image and Vision Computing (IMAVIS), Special Issue on Facial Landmark Localisation "In-The-Wild". 2016.
C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. **300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge**. Proceedings of IEEE Int'l Conf. on Computer Vision (ICCV-W), 300 Faces in-the-Wild Challenge (300-W). Sydney, Australia, December 2013.
C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, M. Pantic. **A semi-automatic methodology for facial landmark annotation**. Proceedings of IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR-W), 5th Workshop on Analysis and Modeling of Faces and Gestures (AMFG 2013). Oregon, USA, June 2013.

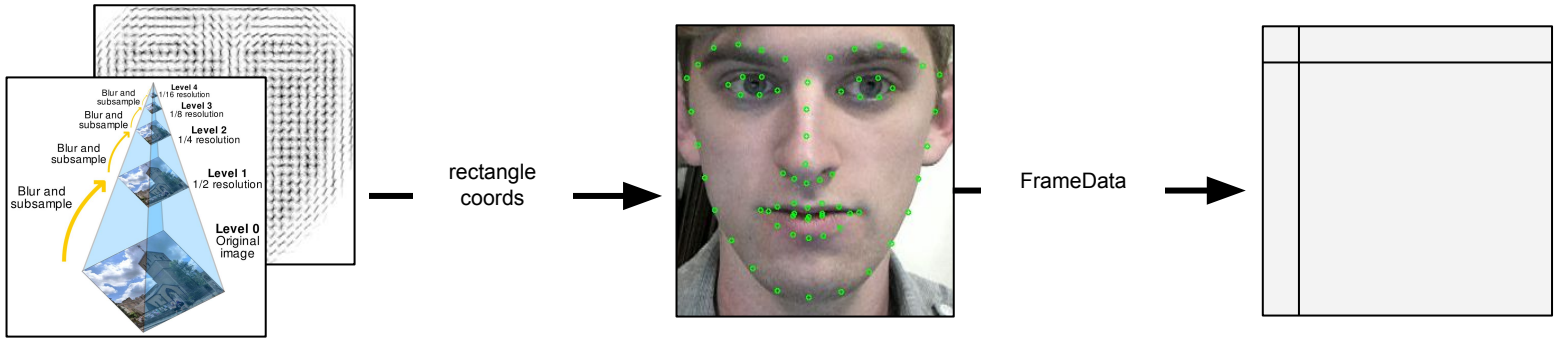
# Linear classifiers




https://en.wikipedia.org/wiki/Pyramid_(image_processing)

# Image pyramids

Sliding window detection scheme




One Millisecond Face Alignment with an Ensemble of Regression Trees by Vahid Kazemi and Josephine Sullivan, CVPR 2014

# Face detection pipeline

# Downsides to my approach

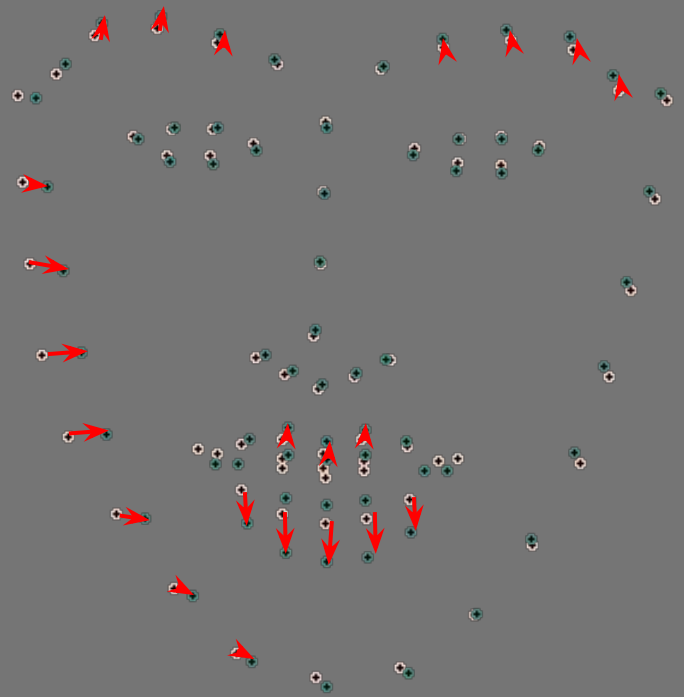
- Single camera = two dimensions
- POSIX Shared Memory compatible only systems

# Current performance

- Pipeline fully functional
- Two-dimensional data
- Localized delta calculations
- Max of 120 fps possible at 640x480!
  - Max 90 fps @ 1280x720
  - Max 60 fps @ 1920x1080
  - ~26 fps @ 4k
- Capture quality needs work

# Localized delta calculations

- Neutral face definition
- Localized face defined by deltas from neutral face
- Pitch, yaw, roll not yet calibrated for



# Demo time!

# Optimizing performance

Current:

- Face detection sampling rate
- Image downsampling

Future:

- Adaptive sampling rate
- Regional face detection
- Multi-threaded and/or GPU-based face detection

# Known issues

- Face identification failure with expressions/gaze too distant from training data
- Lighting conditions must be decent
- Capture system requires consistent face direction (for now)
- Noisy capture data (needs filtering and/or higher resolution image input)
- Definitive bottleneck: **face detection**

# Future goals

- Calibrate face rig data for pitch, yaw, and roll of the head
- Eye tracking
- Cross-platform
- Integration with other animation software
- Endless quality improvement & optimization

# Possible future use cases

- Visdev for 3D animation
- Input to facial expression driven applications (games, interactive media)
- Emotion recognition
- Low bandwidth/high compression video communication system
- Medical examination tool / symptom detection

# Types & applications of computer vision

# Recognitional tasks

- Object recognition/classification
  - Facial detection
- Identification
  - OCR
  - QR codes
  - Facial recognition
- Detection
  - Pose estimation

# Motion analysis

- Tracking
- Egomotion (3D camera tracking)
- Optical flow (apparent motion)

## Other uses

- Scene reconstruction
- Image restoration

# Applications

- Automatic inspection
- Identification tasks
- Detection/surveillance
- Human/computer interaction (input)
- Modelling objects/environments
- Navigation
- Organization

# Questions?

Project homepage:
https://www.sjsuvrlab.org/project_stringless.html

Source code is available!
https://github.com/justint/stringless

http://justintennant.me