Facial appearance scanning using machine vision

Andy Bastable
Lead Engineer, Rare/Microsoft
@andybastable #GDCEurope
Introduction

• **What this talk is**
  • An overview of some of the challenges and general principles when using machine learning in games
  • Inspiration about what’s possible with Kinect
  • Inspiration about what’s possible with machine learning and machine vision

• **What this talk isn’t**
  • A step-by-step guide for writing machine vision systems
  • An academic breakdown of algorithms and code
Champion scanning in Kinect Sports Rivals

• What is the Champion Scan?

• Vision: “Awesome You”
  • Had to be You
  • Had to be Awesome

• Main problems:
  • Would it even work?
  • Making the experience playful yet accurate
  • Knowing that it was going to work for everyone
Primer on Kinect Feeds

- **RGB feed**
  - 1080p
  - 30fps

- **Active IR feed**
  - Lighting independent
  - 30fps

- **Depth feed**
  - 1cm precision at 3m
  - 30fps
How The Scan Works

• User moves into correct positions and performs certain actions
• We scan body & face
• Classifiers determine facial features
• Results used to assemble final character
Tech made up of Machine Vision "Classifiers"

- Face Shape
- Body Size
- Glasses
- Facial Hair
- Skin Tone
- Hair Style
Face Shape

• Developed jointly by Rare and a team from Redmond

• Approach is part of the Xbox One HDFace XDK
  • Available for developers to use
  • Also provides high-quality face animation tracking
  • 10% GPU and 1 Core CPU during shape computation
  • Much less (~12ms) for animation tracking
Face Shape

• First step: register a neutral face mask to the face.
Face Shape

- Then deform mask to “shrink wrap” onto depth feed
Face Shape

• Then recursive PCA to extract blendshape parameters
  • http://en.wikipedia.org/wiki/Principal_component_analysis

• 93 parameters (~sliders) in total
Face Shape

- Overdrive the parameters to give more characterisation
- Apply parameters to “stylised” head
Key Learnings

- Optimise the right thing!
  - Avoid vague goals
  - Test your hypothesis
  - Optimise the correct metrics

- You need lots of data
  - A small amount of data leads to false confidence
Body Size

- First we measure height using Kinect skeleton tracking
- Then torso width and extent using depth and RGB feed
- Then apply to final model
Key Learnings

- Hard to get positive aspirational result
  - Weight often key part of visual identity
  - Weight not often “aspirational”
  - Solution was to find an aesthetic which validates size, but not unflattering.
Glasses & Facial Hair

- Raw classifier developed by team in Redmond
- Available as part of Xbox One Expressions XDK
- Uses ActiveIR for lighting independence
- Random Decision Forest classifier, trained with thousands of images
Glasses & Facial Hair

- Expressions API gives us a point result
  - Noisy
  - We need to average the result over frames and test above a tolerance
  - Initially tweaked by hand, then automated in a script

- Facial Hair
  - Not a binary classification
  - Created a “low confidence” beard to cope with false positives
Key Learnings

- Machine learning does not have to be complex!
  - Can be as simple as a brute force offline tool

- Weight your failure cases to get best results
  - Score “correct” results highest
  - Then “acceptable”
  - Then “ok”
  - And sort results to give the best overall result.
Skin Tone

• Your brain is very good at estimating lighting models.
Computers are not.
Skin Tone

- How important is lighting?
Skin Tone

• The solution: Active IR

• First, we correct for the lighting

Then we average the pixels and compare with known ranges

0.98875 = “light medium” =
Key Learnings

• You need to manage expectations with Machine Learning systems.
  • Perception was that it was “easy” problem
  • Unlikely that any Machine Learning system will hit 100%

• Identify problem data for your approach, and source more of it!
  • We gathered lots of clips of people in low-light conditions
  • Allowed us to quickly test hypotheses to see if they showed promise
  • Iterate, Iterate, Iterate!
Hair Style

- Uses combination of depth feed & hair segmentation in RGB
- Estimates volumes of hair for: Top, Side & Below Ear
- Picks most appropriate hair asset based on results
Hair Style

- If subject is male and has very short hair we run a bespoke baldness classifier
  - Looks for curvature of the forehead
  - Looks to find hairline using RGB feed
Fringes & Pony Tails

• Use hair segmentation for fringe

• If subject is female and has short hair, we assume a pony tail
Hair Colour

• Average RGB pixels from hair
Key Learnings

• It’s OK to cheat
  • If we don’t detect long hair on females we assumed a pony tail
  • Least offensive “wrong” result

• If a result is sensitive add as many layers of security as possible
  • 3 separate tests for baldness
  • Very low false positive result
Animation

• The final result needed to be deformable, yet animate
  • In total 490 blendshapes to deform and animate head
  • Full animation rig mapped onto blendshapes

• GPU bottleneck was transferring blendshapes to GPU
  • Optimisation was to bake “static” blendshapes into new mesh
  • Only transfer animation blendshapes
How We Validated Our Results

• We sourced 850 clips of people being scanned.
• 6 territories (London, Madrid, Turkey, Japan, China, US)
• Strategic mixture of age, gender & ethnicity
How We Validated Our Results

• Each clip annotated to give “ground truth” details about subject.
  • Simple csv file with id, path to recording and results expected

<table>
<thead>
<tr>
<th>clip</th>
<th>age</th>
<th>gender</th>
<th>exp_facialhair</th>
<th>exp_glasses</th>
<th>exp_haircol</th>
<th>exp_haircategory</th>
<th>exp_fringe</th>
<th>exp_skin</th>
<th>exp_body</th>
</tr>
</thead>
<tbody>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>brown</td>
<td>short</td>
<td>yes</td>
<td>light</td>
<td>medium</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>no</td>
<td>yes/no</td>
<td>no</td>
<td>grey</td>
<td>short</td>
<td>no</td>
<td>light</td>
<td>thin</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>blonde/brown</td>
<td>short/buzz</td>
<td>no</td>
<td>light</td>
<td>thin</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>brown</td>
<td>short</td>
<td>no</td>
<td>light</td>
<td>large</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>black</td>
<td>short</td>
<td>no</td>
<td>light</td>
<td>thin</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>blonde</td>
<td>short</td>
<td>yes</td>
<td>light</td>
<td>thin</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>yes/no</td>
<td>no</td>
<td>yes</td>
<td>bald</td>
<td>bald</td>
<td>no</td>
<td>light</td>
<td>medium/large</td>
</tr>
<tr>
<td>rexaadult</td>
<td>female</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>brown</td>
<td>pony</td>
<td>no</td>
<td>light</td>
<td>medium/large</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>yes/no</td>
<td>no</td>
<td>no</td>
<td>brown</td>
<td>short/tall</td>
<td>no</td>
<td>light</td>
<td>thin</td>
</tr>
<tr>
<td>rexaadult</td>
<td>female</td>
<td>yes/no</td>
<td>no</td>
<td>no</td>
<td>brown</td>
<td>long</td>
<td>no</td>
<td>light</td>
<td>medium/thin</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>black</td>
<td>short/tall</td>
<td>no</td>
<td>light</td>
<td>thin</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>yes/no</td>
<td>no</td>
<td>no</td>
<td>black</td>
<td>short</td>
<td>no</td>
<td>light</td>
<td>thin</td>
</tr>
<tr>
<td>rexaadult</td>
<td>male</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>buzz</td>
<td>no</td>
<td>no</td>
<td>light</td>
<td>thin</td>
</tr>
</tbody>
</table>
How We Validated Our Results

• Automatic process to run each clip with latest code
  • Hooked into our automatic build process
  • Ran on 16 devkits in 3 hours
  • Twice daily
How We Validated Our Results

- Generated html report with full data and deltas

<table>
<thead>
<tr>
<th>Category</th>
<th>Accuracy</th>
<th>Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body Picker</td>
<td>98.8%</td>
<td>-0.2%</td>
</tr>
<tr>
<td>Facial Hair Picker</td>
<td>96.9%</td>
<td>+0.3%</td>
</tr>
<tr>
<td>Adult Males Correct</td>
<td>92.2%</td>
<td></td>
</tr>
<tr>
<td>Glasses Picker</td>
<td>97.6%</td>
<td>-0.6%</td>
</tr>
<tr>
<td>Hair Colour Picker</td>
<td>91.0%</td>
<td>-0.2%</td>
</tr>
<tr>
<td>Hair Style</td>
<td>93.3%</td>
<td>-0.1%</td>
</tr>
<tr>
<td>Hair Length Correct</td>
<td>98.8%</td>
<td></td>
</tr>
<tr>
<td>Baldness Correct</td>
<td>90.3%</td>
<td></td>
</tr>
<tr>
<td>Prime Correct</td>
<td>94.5%</td>
<td></td>
</tr>
<tr>
<td>Skin Colour Picker</td>
<td>95.4%</td>
<td>+0.0%</td>
</tr>
<tr>
<td>Light Skin Correct</td>
<td>99.7%</td>
<td></td>
</tr>
<tr>
<td>Medium Skin Correct</td>
<td>95.0%</td>
<td></td>
</tr>
<tr>
<td>Dark Skin Correct</td>
<td>92.4%</td>
<td></td>
</tr>
</tbody>
</table>
How We Validated Our Results

- We were able to track progress over time
Key Learnings

• Machine learning lives or dies on the quality of source data

• 24 hour cycle of improve, observe, validate, repeat

• Cut corners where you can
  • You are unlikely to hit 100%, so goal is to maximise results
  • Test a simple assumption, it might save a lot of work
Some Results
Some Results
Some Results
Some Results
Some Results
Some Results
Scanning Experience

- Whimsical/playful tone
- Dr Who!
- Required a LOT of User Research
- Biggest challenge: positioning the user
Scanning Experience

• 24 hour cycle of User Research and reaction
  • All engineers observed sessions
  • Quick deadline to verify improvements
Scanning Experience

• The reveal cutscene
  • Create tension and anticipation
  • Fun payoff to the experience
What Went Well?

• End result is good
• Scanning flow works well for almost all users
• Machine vision works well for most users
• Automated testing gave us launch confidence
What Could We Have Done Better?

• Data Capture was started late.
  • Get data early!

• “Experience” user research was started early enough but not initially useful due to missing build functionality

• Result trends towards generic for ~50% of users

• Hair styles were correct, but often uninspiring
Any Questions?

andy.bastable@microsoft.com
@andybastable
#GDCEurope