Performance-driven Facial Animation

Garoe Dorta Perez, Ieva Kazlauskaite, Richard Shaw

University of Bath
Centre For Digital Entertainment

27 May 2015

Outline

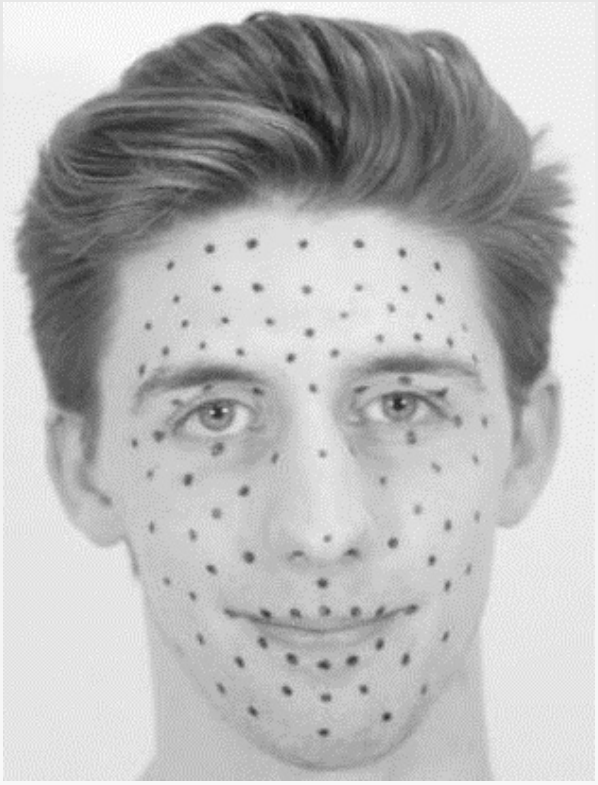
Data Capture

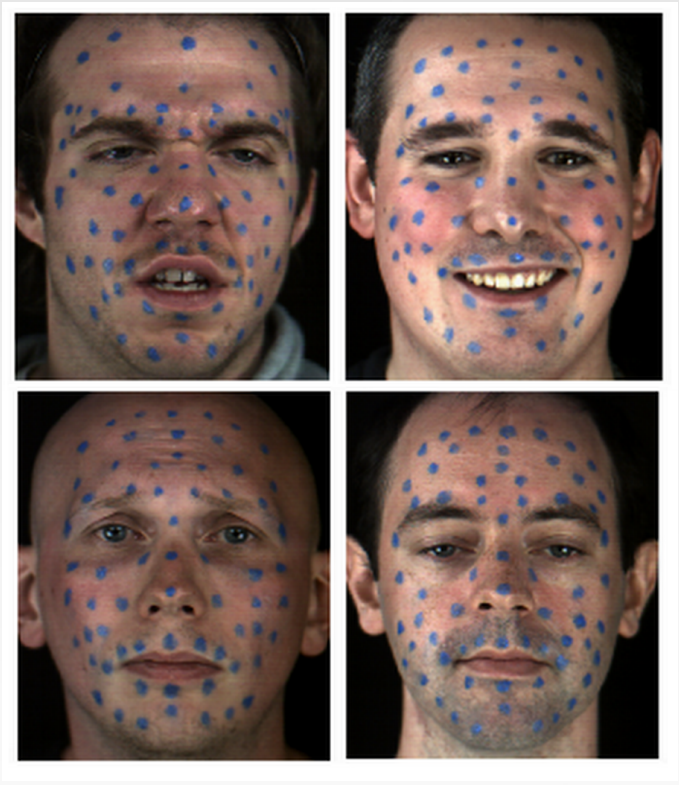
- Performance capture using two DSLR cameras in stereo.
- Video recorded at 60fps with resolution 640×480 pixels.
- Video streams synchronised using audio signals.

The data capture session using two DSLR cameras in stereo.

Face Markers

- Markers are drawn onto the actor's face to track the facial performance.

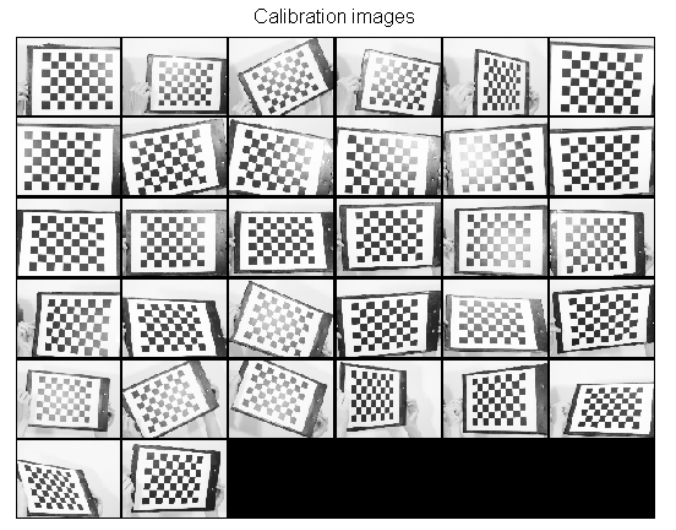


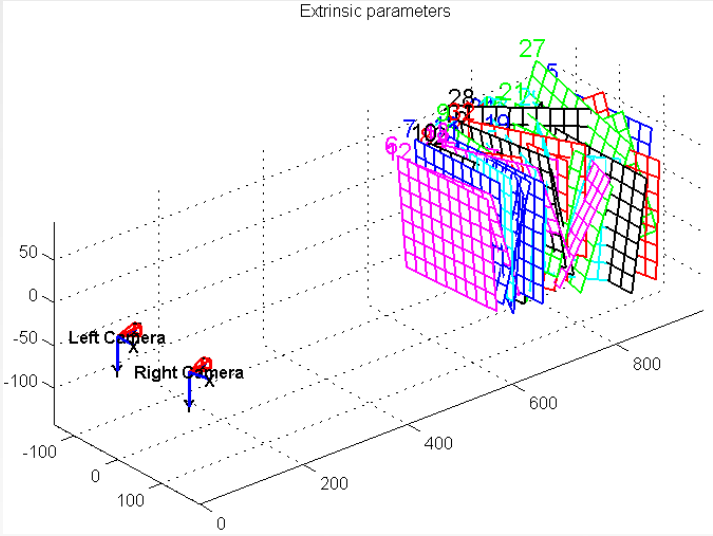


Marker positions were roughly based on the Surrey Audio-Visual Expressed Emotion (SAVEE) Database.

Stereo Calibration

- A checkerboard pattern used to calibrate stereo camera setup.
- Obtain the cameras' intrinsic and external parameters.
- Compute the projection matrices and fundamental matrix.





The stereo cameras were calibrated using a checkerboard pattern.

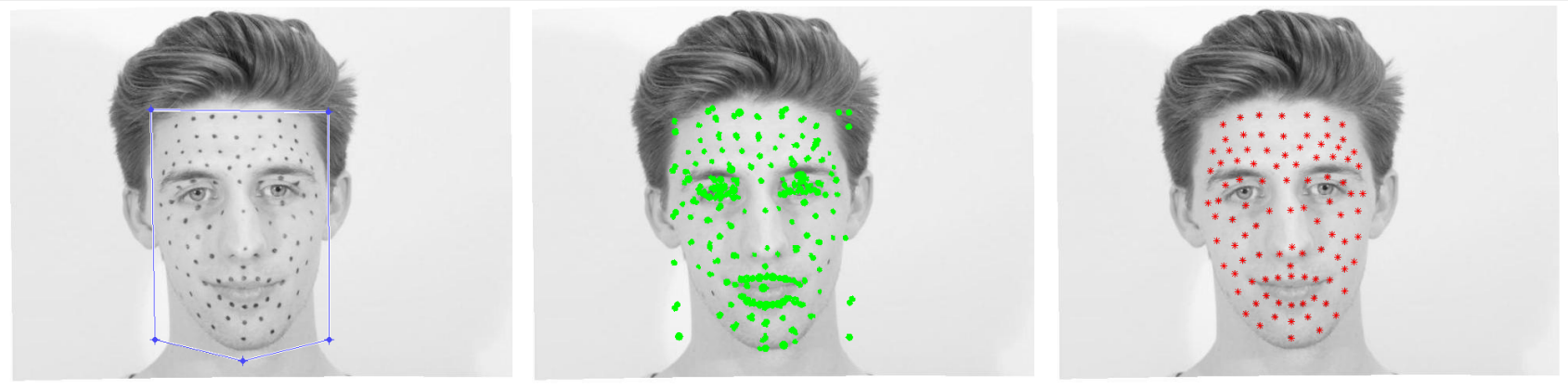
Stereo Rectification

- Using the camera parameters we rectify each stereo pair for a captured image sequence.
- Corresponding epipolar lines lie on the same pixel rows.
- Reduces correspondence problem to a 1D search.

Each frame of an image sequence is stereo rectified.

Initial Marker Detection

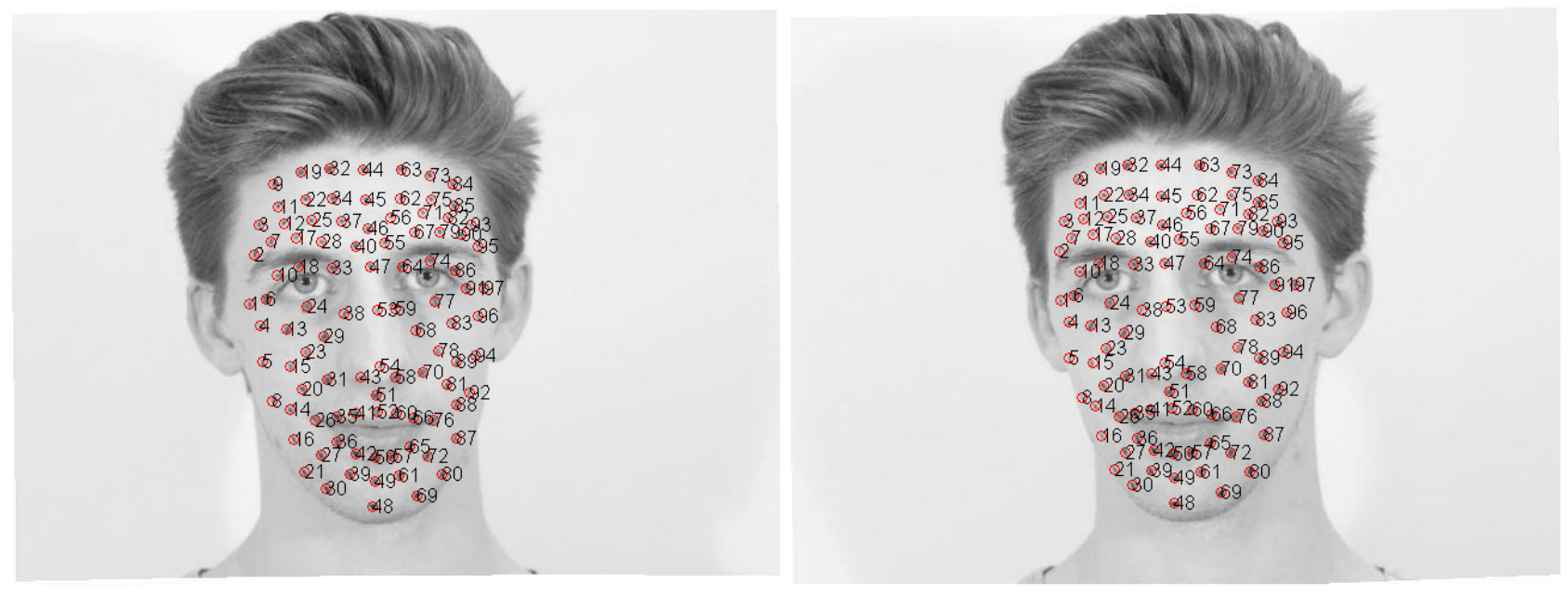
- Face markers are detected by computing SIFT features in the left image of the first frame.
- Wrongly detected pixels can be removed and additional points included interactively.



Markers on the face are detected using SIFT features.

Feature Matching

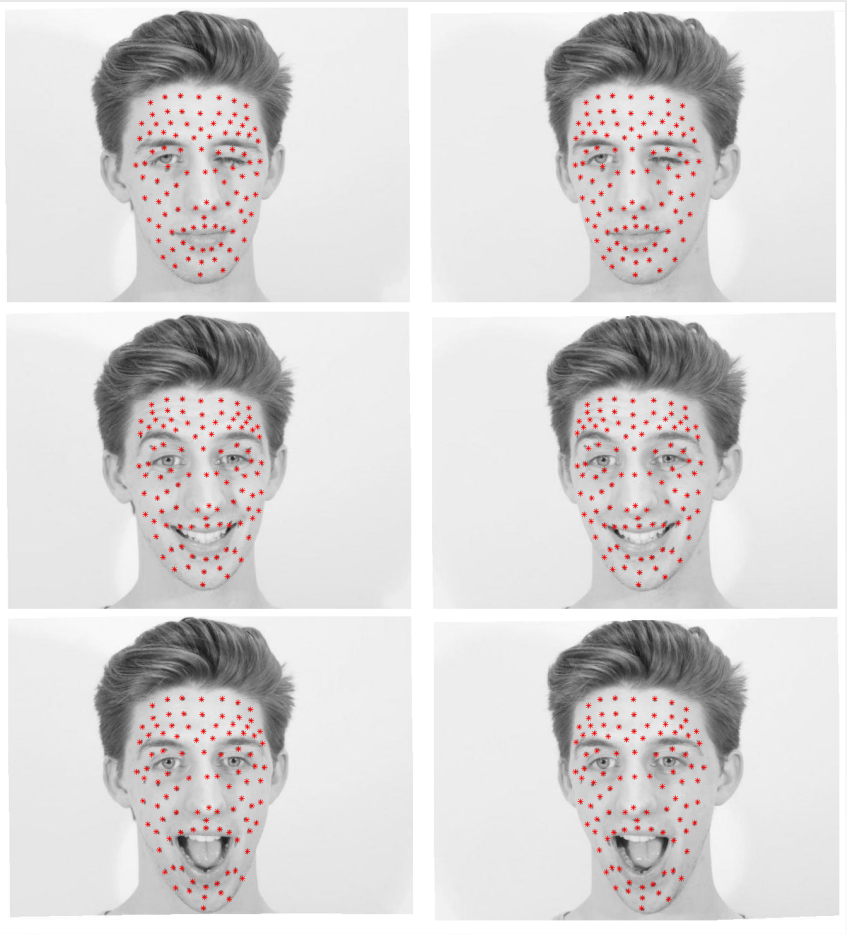
- Corresponding markers in the right image are found by searching along the corresponding epipolar lines.
- Matching features are found by computing the normalised cross-correlation for image patches around each feature point.



Corresponding markers are found in the right image.

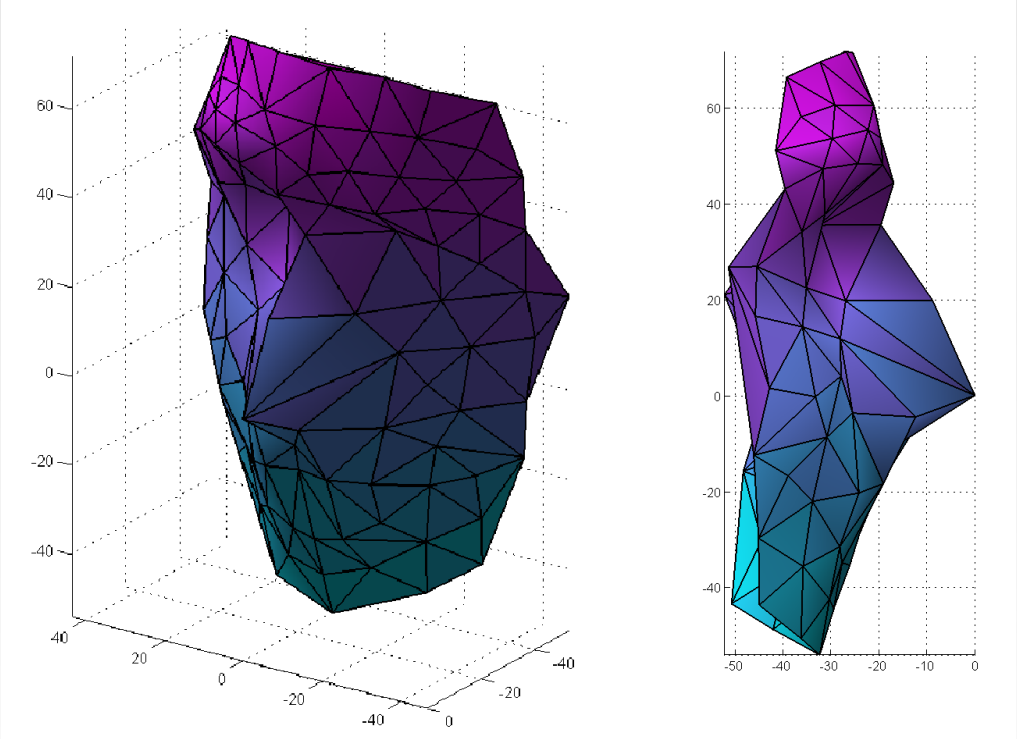
Marker Tracking

- We track the face markers throughout the entire image sequence using the KLT tracking algorithm.



Sparse 3D Reconstruction

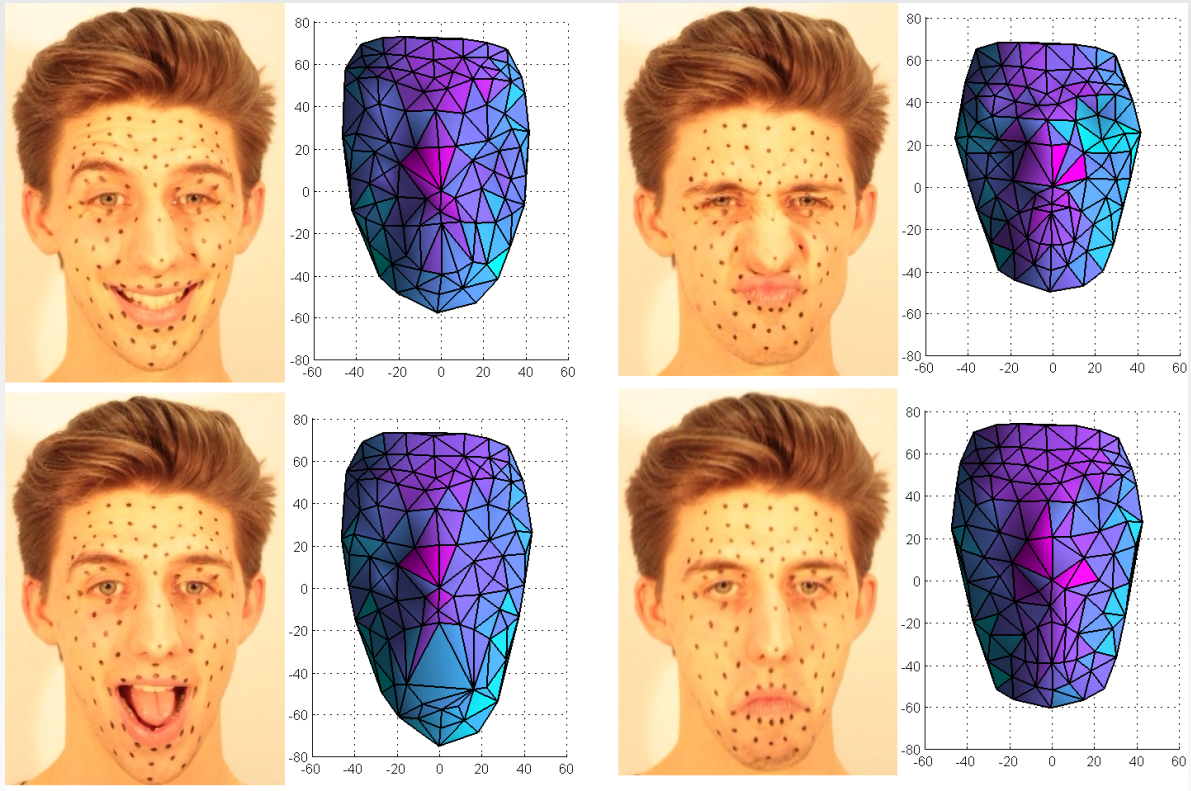
- Using the known camera projection matrices, we compute a sparse 3D reconstruction by triangulating.



The 3D reconstruction of the neutral expression.

Sparse Sequence Reconstruction

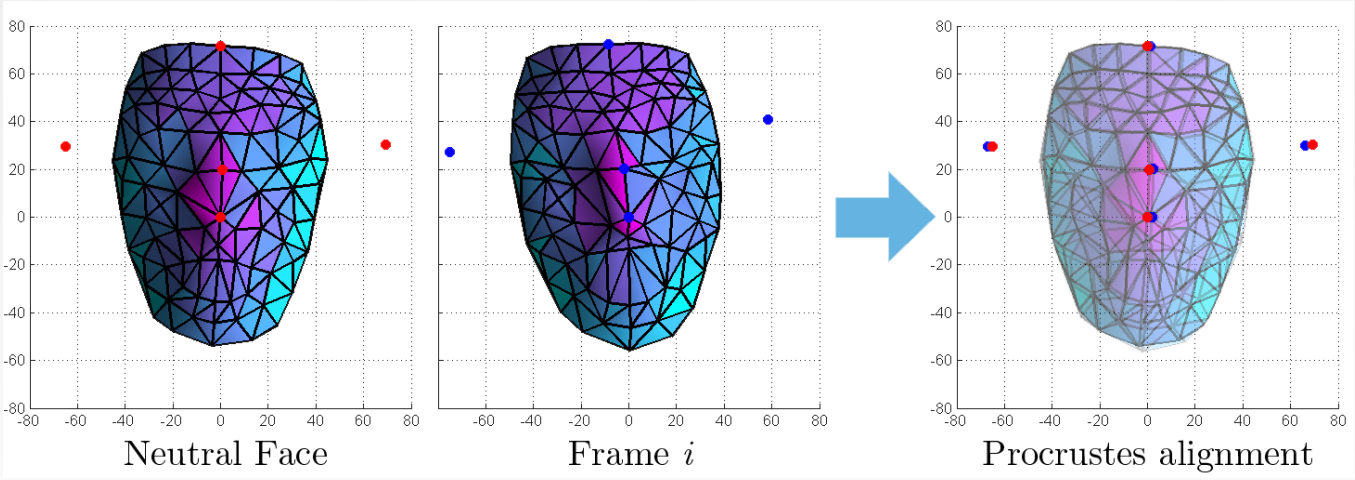
- Using the known camera projection matrices, we compute a sparse 3D reconstruction for every frame.



A sample of reconstructed frames from a captured image sequence.

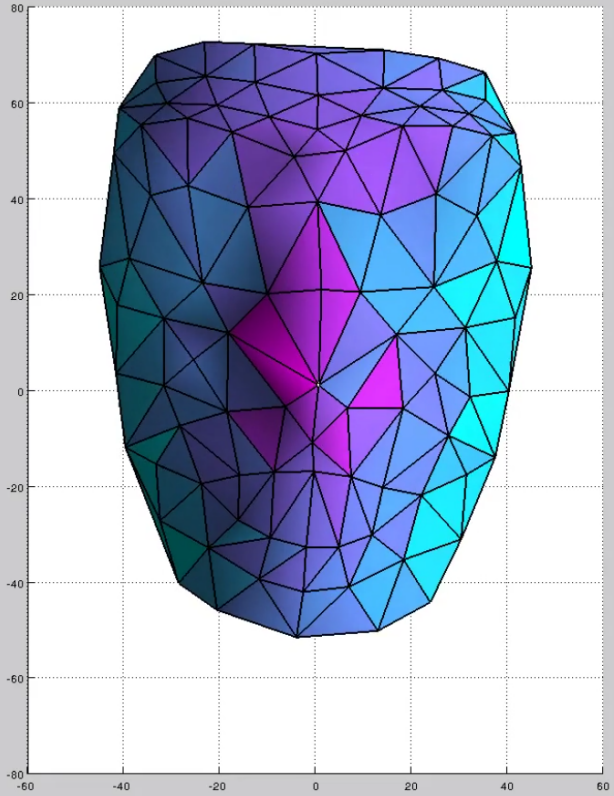
Head Stabilisation

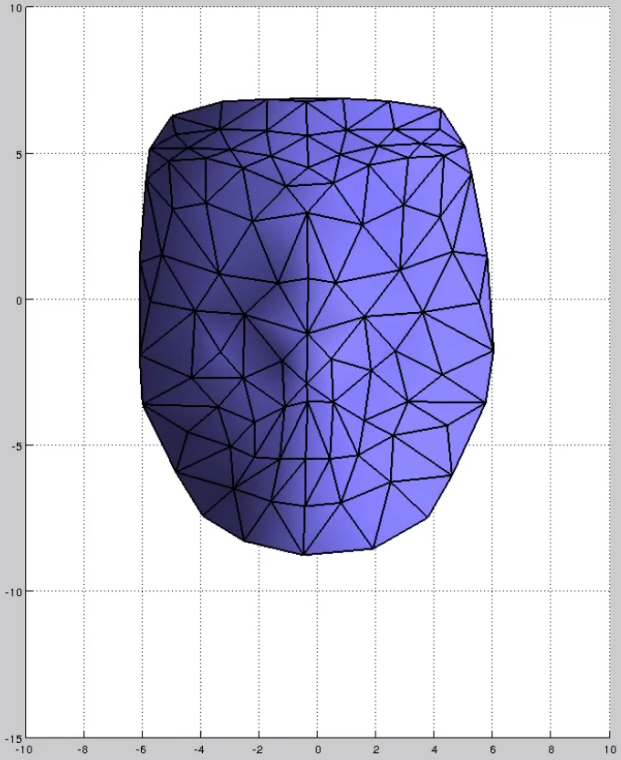
- We attempt to remove the rigid head motion by performing Procrustes alignment.
- The head pose in each frame of an image sequence is aligned to the neutral pose.
- The 3D point trajectories are also smoothed to remove jittery head motion.



Rigid head motion is removed using Procrustes analysis.

Sparse Emily sequence





Outline

Data Capture and 3D Reconstruction

Blendshape Model

Skin Rendering

General idea

Given a set of blendshapes $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$ and a neutral face $\mathbf{b}_0$, a new facial expression $\mathbf{F}(\mathbf{w})$ is a linear combination of the offsets of the basic shapes:

$$\underbrace{\mathbf{F}(\mathbf{w})}_{\text{new expression}} = \underbrace{\mathbf{b}_0}_{\text{neutral face}} + \sum_{i=1}^{N} \underbrace{w_i}_{\text{weights}} \underbrace{|\mathbf{b}_i - \mathbf{b}_0|}_{\text{offsets from neutral}} .$$

The **weights** describe how much each of the basic shapes affect the new expression.

Designing blendshapes

Option 1 Create blendshapes **by hand** $\Rightarrow$ correspond to a basic facial expression, for example a raised eyebrow.
$\times$ Hard to produce.

Option 2 Use **P**rincipal **C**omponent **A**nalysis $\Rightarrow$ automatic construction.
$\times$ Hard to adjust manually.

Estimating weights

The weights $\mathbf{w}$ are estimating by solving the following:

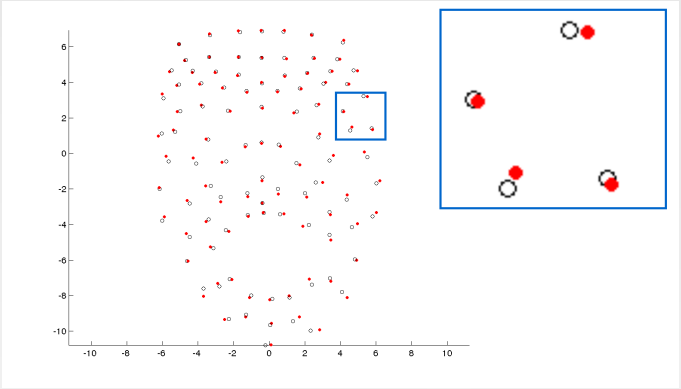
Minimisation problem

$$\min_{\mathbf{w}} \| \underbrace{\hat{\mathbf{B}}}_{\text{matrix with offsets}} \underbrace{\mathbf{w}}_{\text{vector of weights}} - \underbrace{(\mathbf{F}(\mathbf{w}) - \mathbf{b}_0)}_{\text{offsets of new expression}} \|,$$

where the weights add up to 1.

Experiments

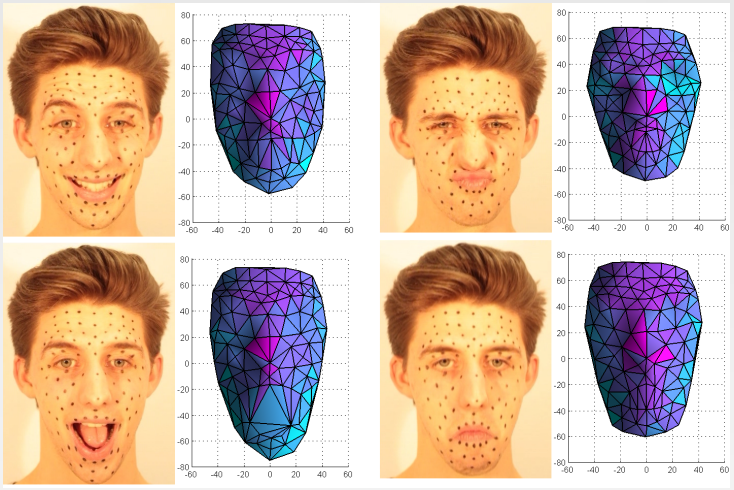
- Performance of numerical methods is compared by measuring **mean squared reconstruction error**.



- Extra shapes.

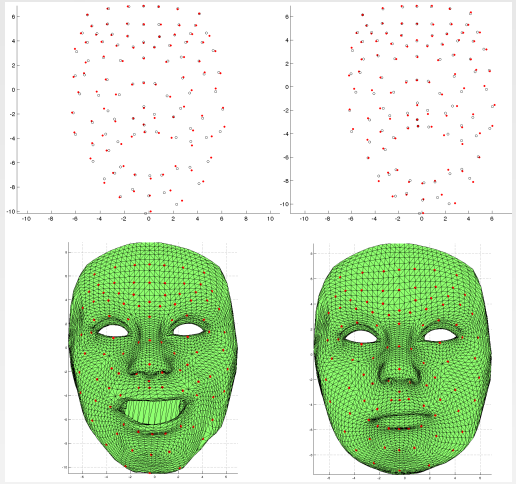
- Different upper bound on each dimension of **w**.

Initial Approach



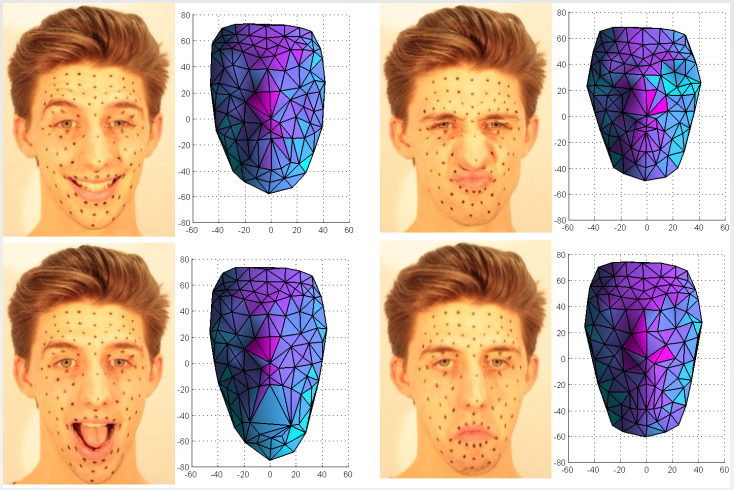
Recorded sequence

Thin plate splines



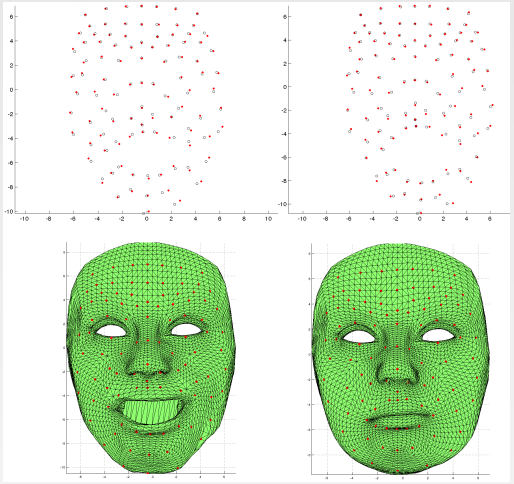
Emily sequence

Initial Approach



Recorded sequence

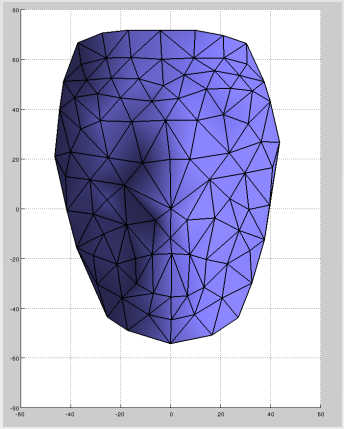
Thin plate splines



Emily sequence

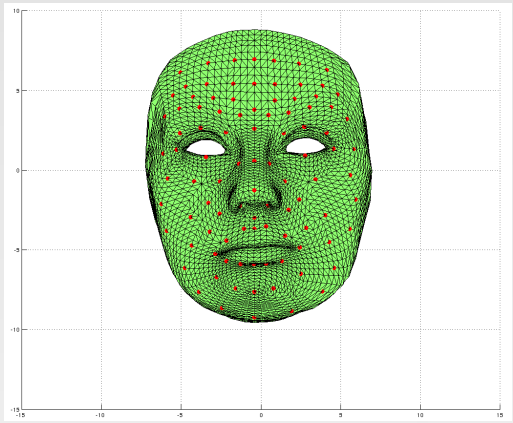
$\Rightarrow$ Solve for weights in Emily domain.

Solve for weights in actor's domain



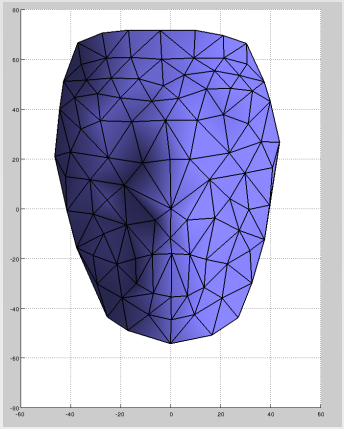
Neutral actor

T



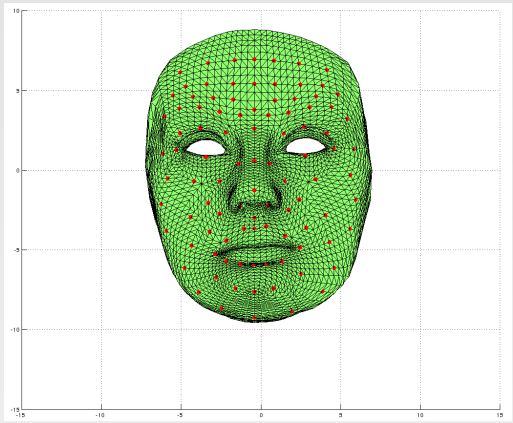
Neutral Emily

Solve for weights in actor's domain

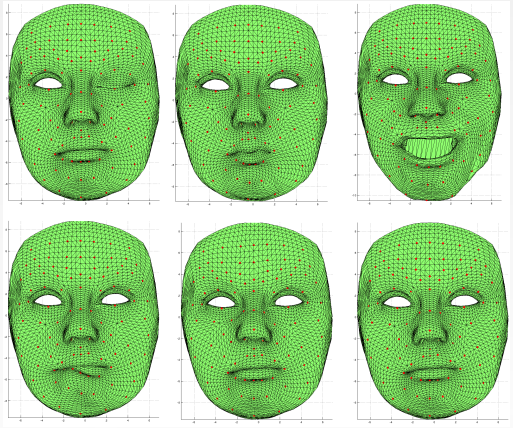


Neutral actor

T

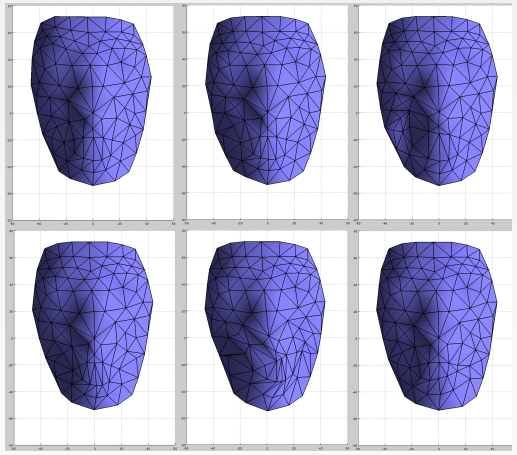


Neutral Emily



Emily blendshapes

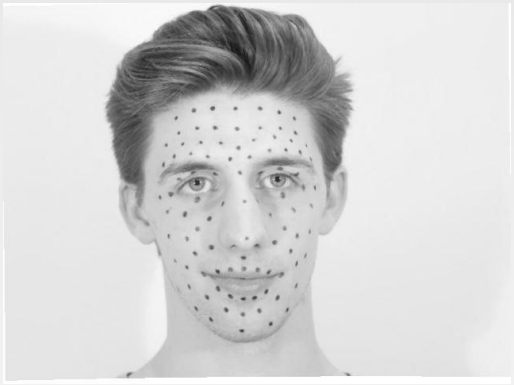
$\mathbf{T}^{-1}$



Actor's blendshapes

Issues

- Mismatch of neutral expressions.



- Lack of expressiveness in Emily blendshapes.
- Lack of expressiveness in the sparse shapes, and sequence.

Outline

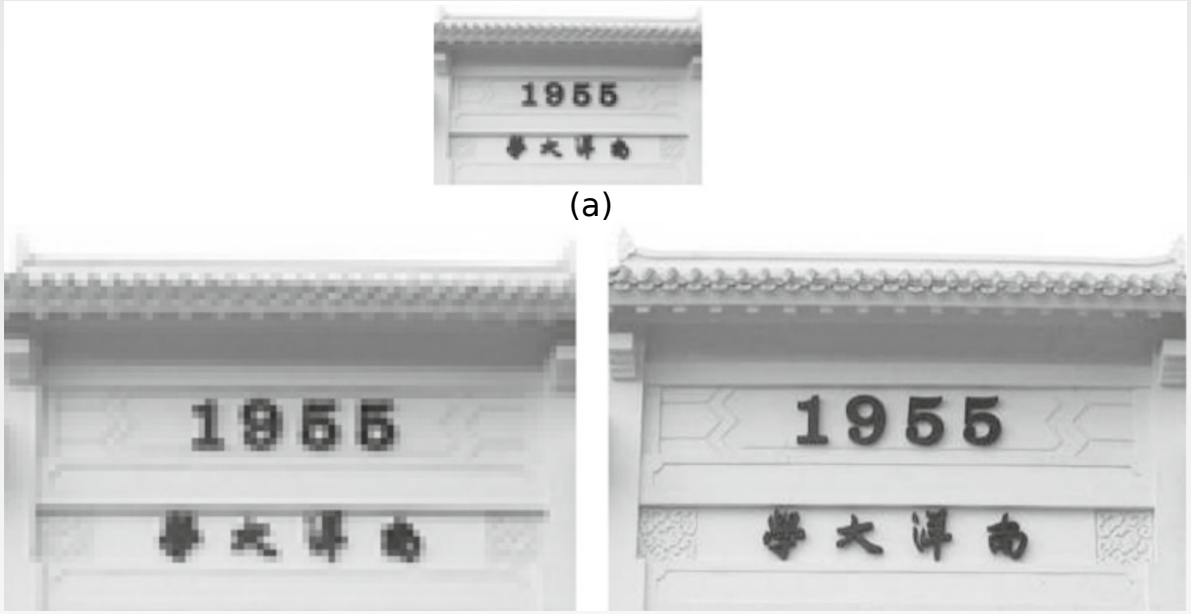
Data Capture and 3D Reconstruction

Blendshape Model

Skin Rendering

Texture super-resolution

- Improve rendering quality by increasing skin texture quality
- Increase the resolution of a texture image



Super-resolution example, (a) is original image, (b) is enlarged with pixel duplication and (c) is enlarge with super-resolution technique.

Image Analogies

- General filter based on image examples
  - Pixel match for A'
  - Coherence match for patch in B'

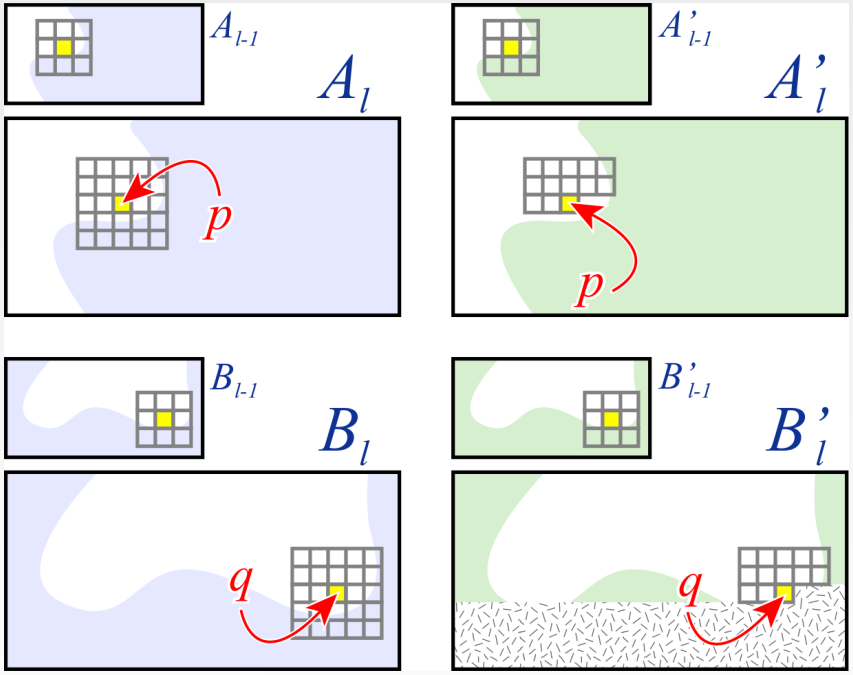
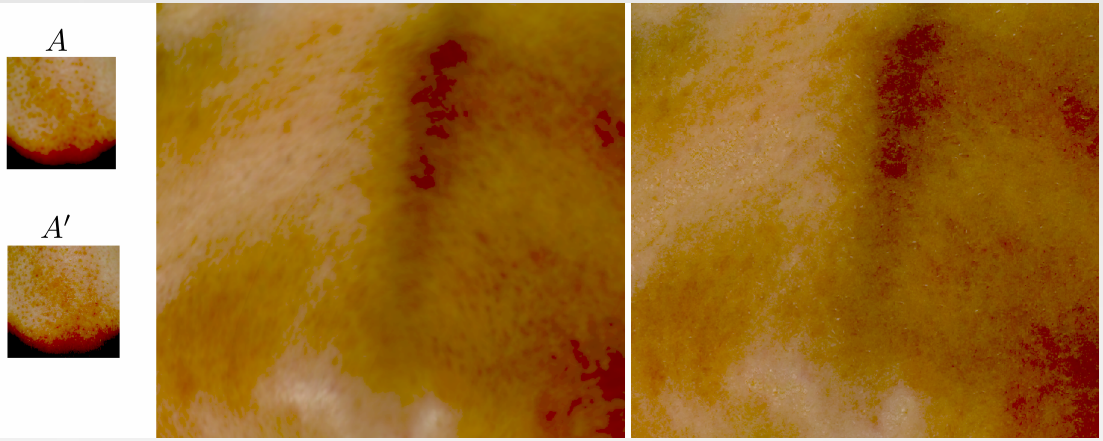


Image analogies diagram, A and A' are example images, B is input image, B' is output image, l is the synthesis level, q is the current pixel in B, B' and p the current pixel in A, A'.

Image Analogies



With inputs A, A' and B, Image Analogies filter outputs B', images is falsed-coloured to highlight differences.

Areas of face segmentation.

Image Analogies

Original texture.

Filtered texture.

Super-resolution with sparse reconstruction

- Built dictionary of high resolution samples
- Synthesize new image taking patches from the dictionary

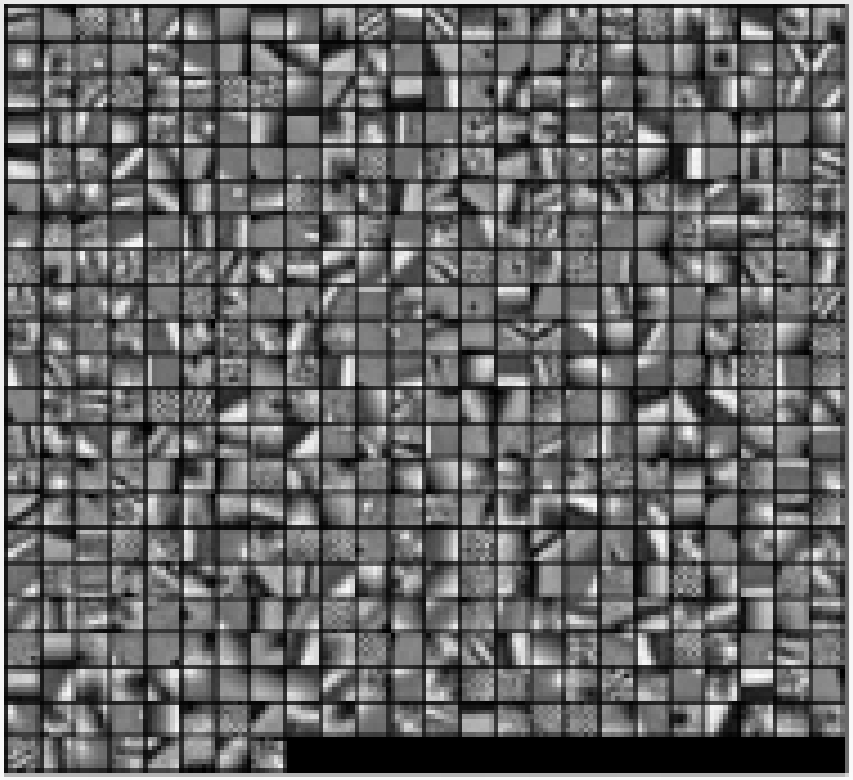
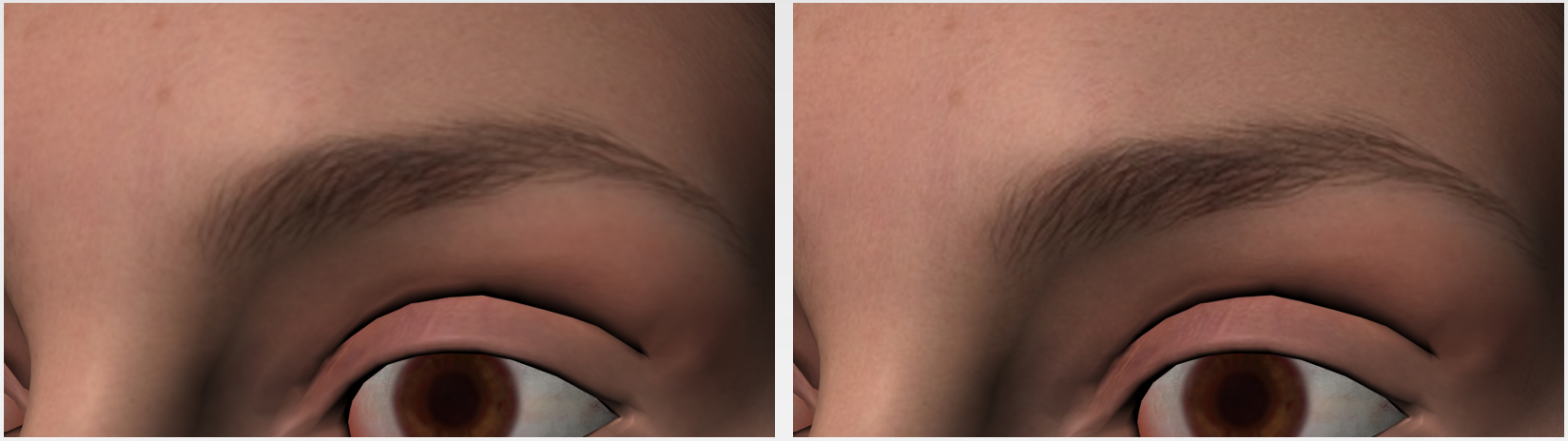


Image patches in the dictionary.

Super-resolution with sparse reconstruction

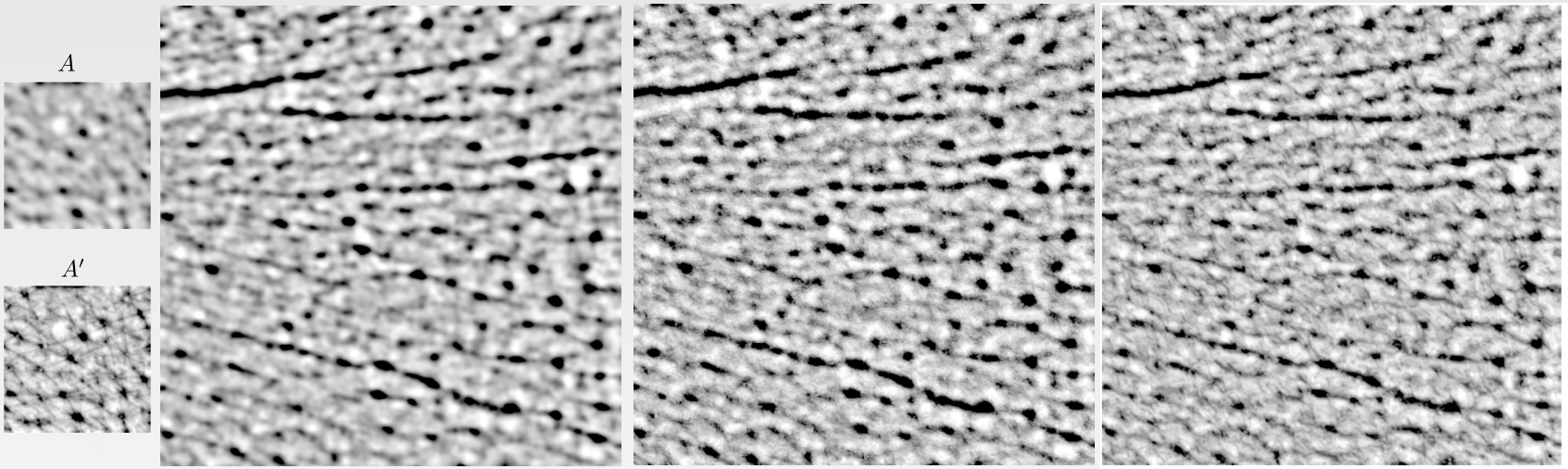
Texture super-resolution for Emily, (left) original texture, (right) synthesized texture with higher resolution.

Super-resolution with sparse reconstruction



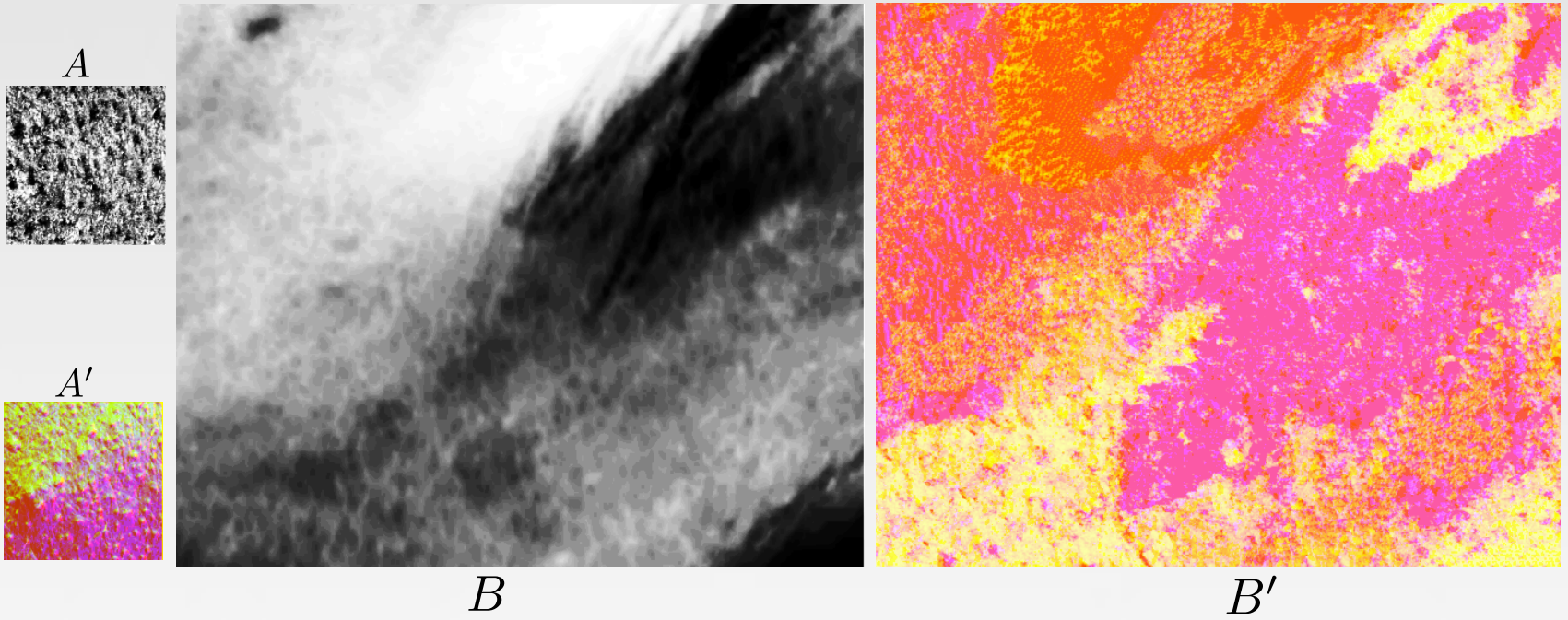
Close up of texture super-resolution for Emily, (left) original texture, (right) synthesized texture with higher resolution.

Image Analogies: Other applications



Adding high frequency detail to a bump map.

Image Analogies: Other applications



Generating normal maps from gray scale textures, images is falsed-coloured to highlight differences.