Mixtures of Eigenfeatures for Real–Time Structure from Texture

Tony Jebara
Kenneth Russell
Alex Pentland

Vision & Modeling
Perceptual Computing
M.I.T. Media Laboratory

MOTIVATION

– Recover full 3D information from live video:
rotation, translation, focal length, structure, texture
– Structure is underconstrained from a single frame
usually constraints / regularization compensate:
a) local smoothness (shape from shading)
b) motion (structure from motion)
c) multiple images (stereo techniques)
d) knowledge of a class of objects
e) global vs. local structure constraints

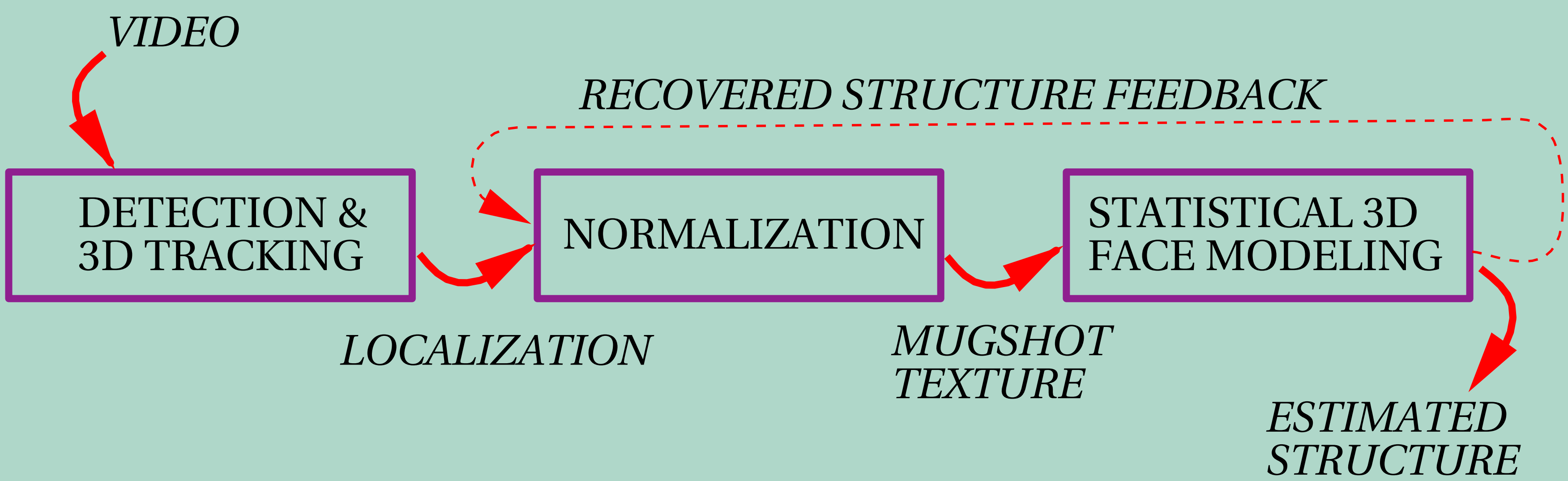
– [CVPR 97]: *Real–time feedback 3D tracker with SfM*
SfM only recovers sparse point field (40 points @ 30Hz)
(unless many features or optical flow –> Not Real Time)

– How much structure info in a static frame?
– Real time structure estimation –> temporal analysis
– Fully reconstruct 3D face compactly (video–conf)
– Can use global linear analysis vs. local non–linear
redundancies and regularities in global data

SYSTEM ARCHITECTURE

– Learn the mapping between texture and structure probabilistically to avoid cumbersome low–level operations and underconstrained formulations

– Use pre–processing and tracking to focus learning resources where needed.



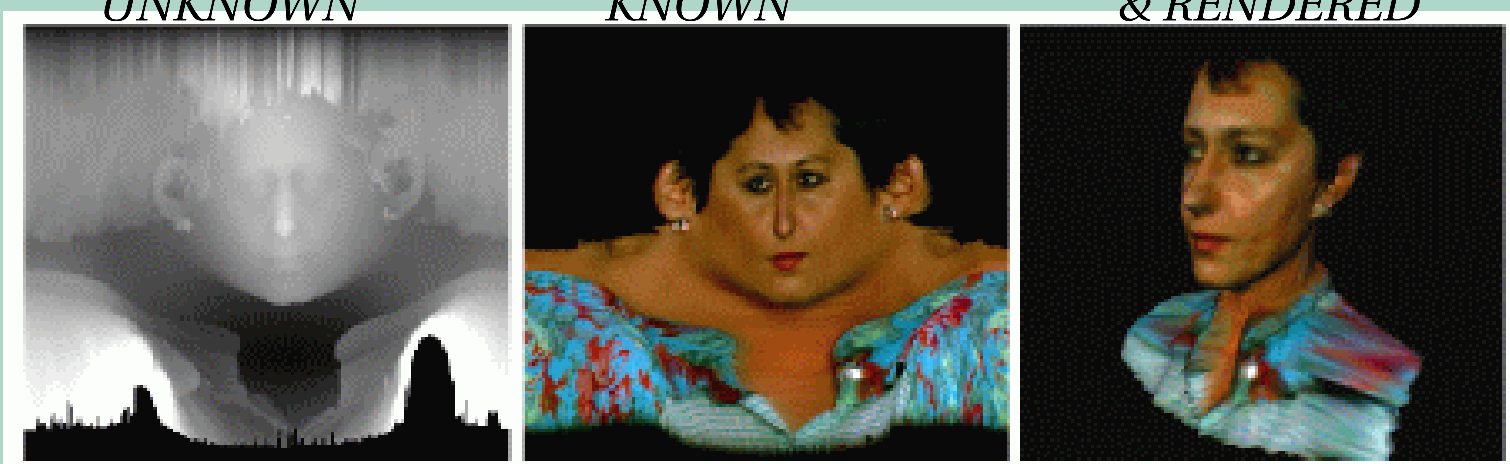
–Want a compact real–time temporal representation of facial 3D pose, translation, structure and texture

3D TRAINING DATA

–Knowledge Based Statistical Approach
Learn mapping between texture and structure from Cyberware 3D face scans.

–360 Degree Cylindrical (R,G,B,radius) value for each (theta,y) value. (very large dimensionality)

–Database of over 300 Scans:
Wright Patterson Air Force, Lab Personnel,Headus (Cyberware affiliate), Other...



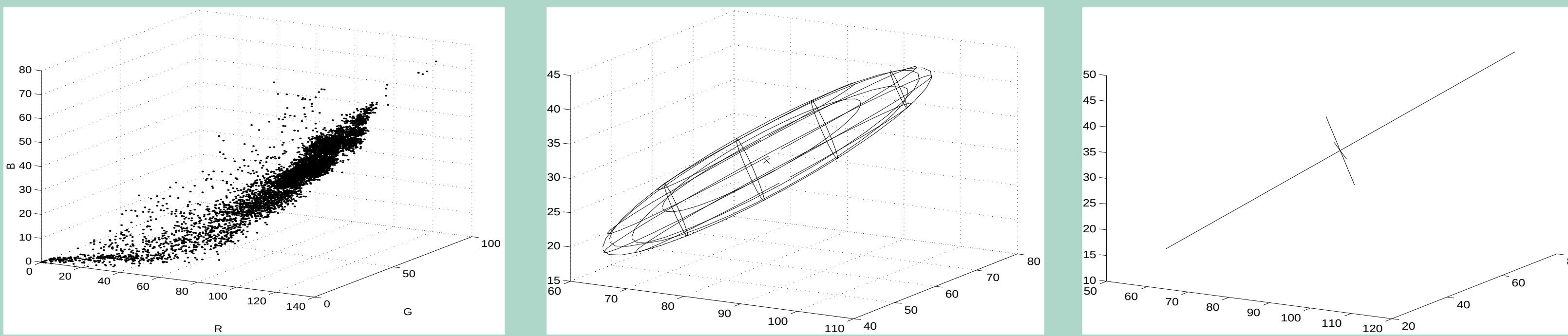
PRE–PROCESSING

– Remove unwanted variations and non–linearities (such as illumination, background and 3D pose) to uncover interesting correlations between texture & structure

– Segmentation and 3D alignment (semi–rigid transformation)

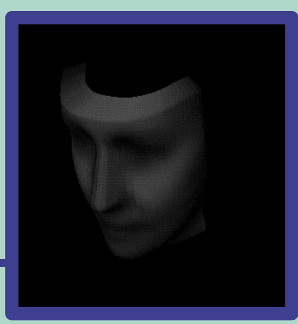
–Illumination Correction (RGB not independent)

Eigenspace Basis over facial colors for histogramming



Histogram fitting via cdf’s

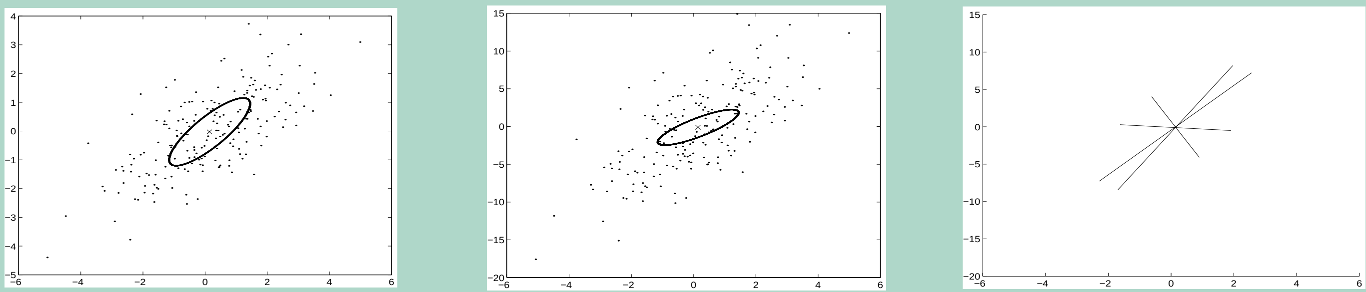
$$T_{s->d}(i) = c_d^{-1}(c_s(i))$$



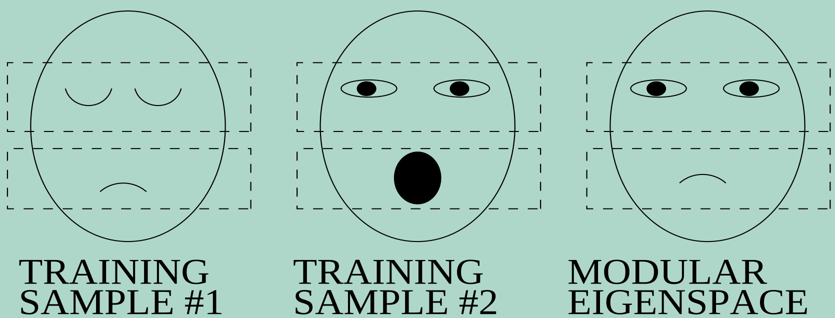
EIGENFEATURES

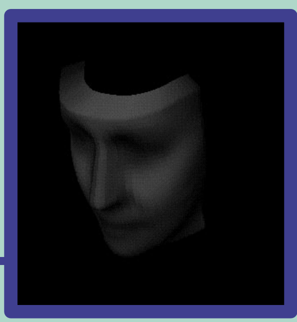
– Form a rank–defficient Gaussian model over the data (of dims M >> N samples)

– PCA sensitive to scaling (but not rotation):



–Focus PCA resources where needed using modular eigenspace with a weight for each pixel





–Estimate several smaller M using N samples
–Span a larger space

MIXTURES OF EIGENFEATURES

– Decouple image spatially into multiple eigenspaces each with its own region of applicability (weight) forms a ***mask*** for each eigenspace

–Make a more generalizable space than full PCA

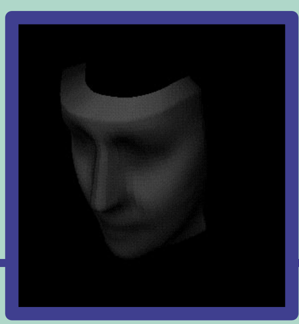
– Simple iterative algorithm based on assumption that the Modules are binary and don’t overlap

Inject regularizers (blur, epsilon, symmetry)

Estimate eigenspaces over training set

Compute eigenspaces error over x–validation set

Partition image into modules according to errors

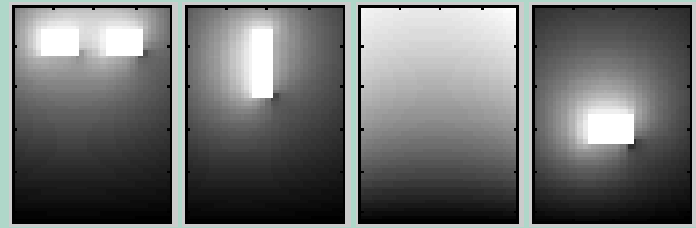


– Initialize the model pseudo–randomly and iterate until convergence (min. cross–validation error)

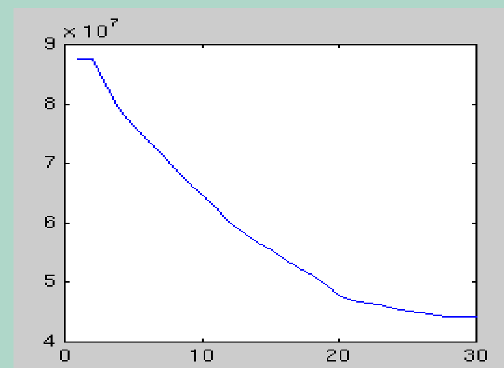
FINDING THE EIGENFEATURES

– Decouples face into multiple regions with their own linear models and modes.

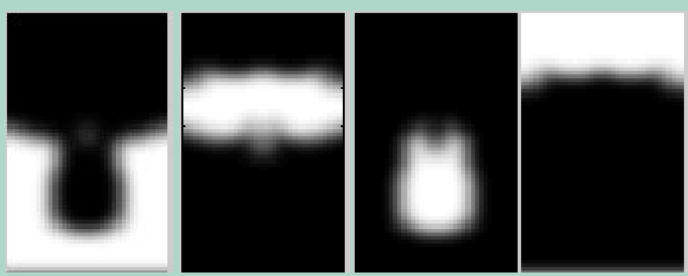
– Initialized both *randomly* and *manually* using Laplace’s equation for interpolation

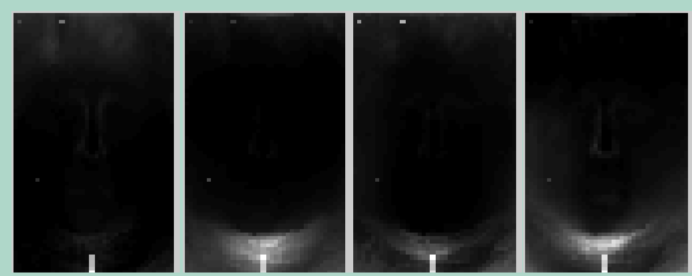


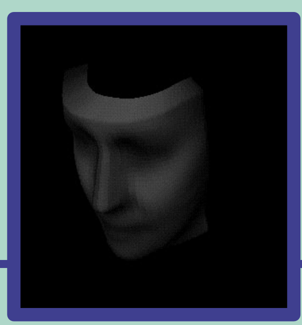
– Converges to a local min error, reduces Error up to 70% after 25 iterations



– Resulting Masks & Spatial Errors

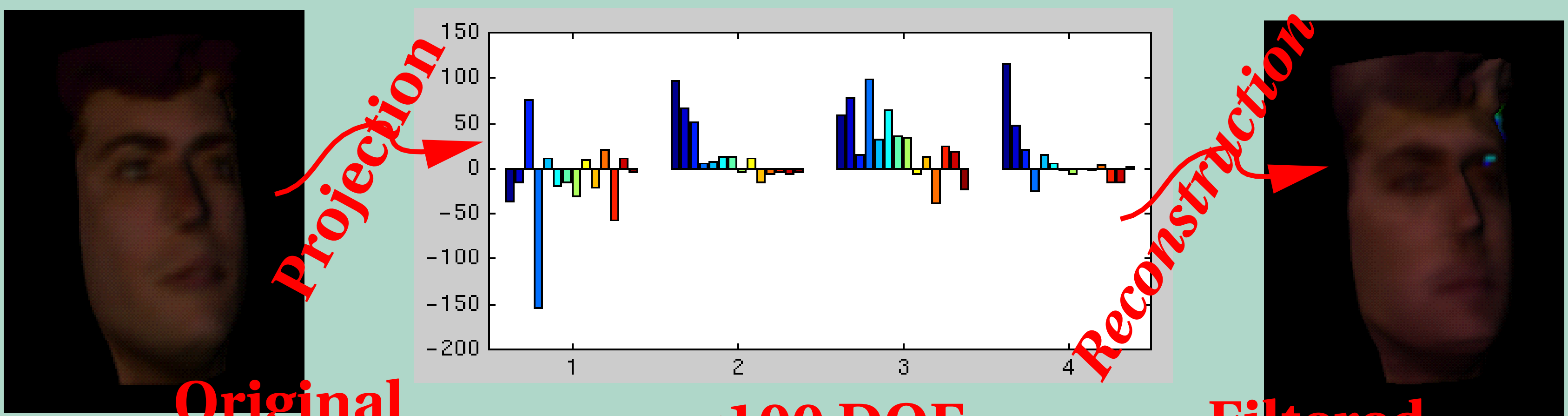






EIGENVECTORS, PROJECTION & FILTERING

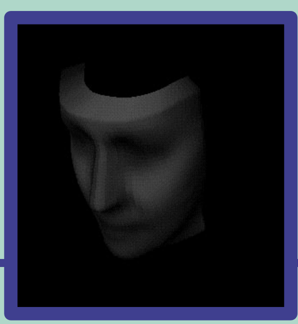
– Modular Eigenvectors resulting from training show coupling over image regions



Original
10,000+ DOF

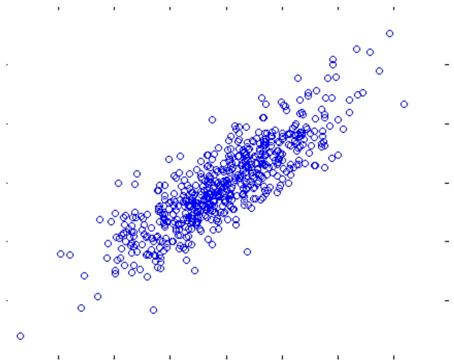
<100 DOF

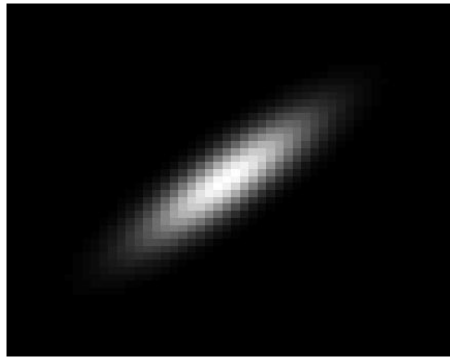
Filtered
10,000+ DOF

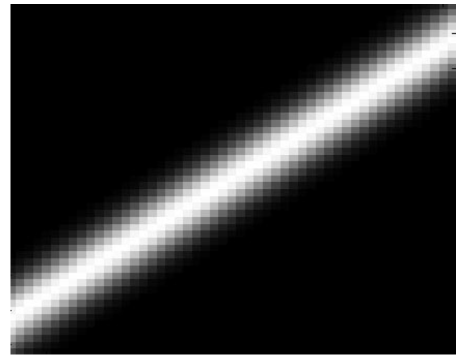


Compact Coefficient Representation
4x14 Floats, storage size < 100 bytes

REGRESSION





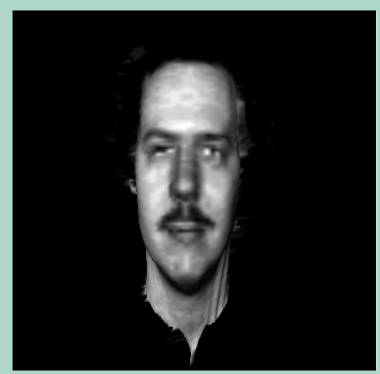


$$x^* = \begin{bmatrix} x_k \\ \hat{x}_u \end{bmatrix} = \begin{bmatrix} x_k \\ \Sigma_{uk}\Sigma_{kk}^{-1} x_k \end{bmatrix}$$

$$p(x_k, x_u) = \frac{exp\{-\frac{1}{2}\begin{bmatrix} x_k \\ x_u \end{bmatrix}^T \begin{bmatrix} \Sigma_{kk}\Sigma_{uk} \\ \Sigma_{uk}\Sigma_{uu} \end{bmatrix}^{-1} \begin{bmatrix} x_k \\ x_u \end{bmatrix}\}}{(2\pi)^{D/2}\sqrt{\begin{vmatrix} \Sigma_{kk}\Sigma_{uk} \\ \Sigma_{uk}\Sigma_{uu} \end{vmatrix}}}$$

$$p(x_u|x_k) \propto exp\{ -\frac{1}{2}(x_u - \Sigma_{uk}\Sigma_{kk}^{-1}x_k)^T \times (\Sigma_{uu} - \Sigma_{uk}\Sigma_{kk}^{-1}\Sigma_{ku})^{-1} \times (x_u - \Sigma_{uk}\Sigma_{kk}^{-1}x_k) \}$$

– Tricky in higher dims due to rank defficiency and cost
– Use trained model,condition on known & estimate unknown

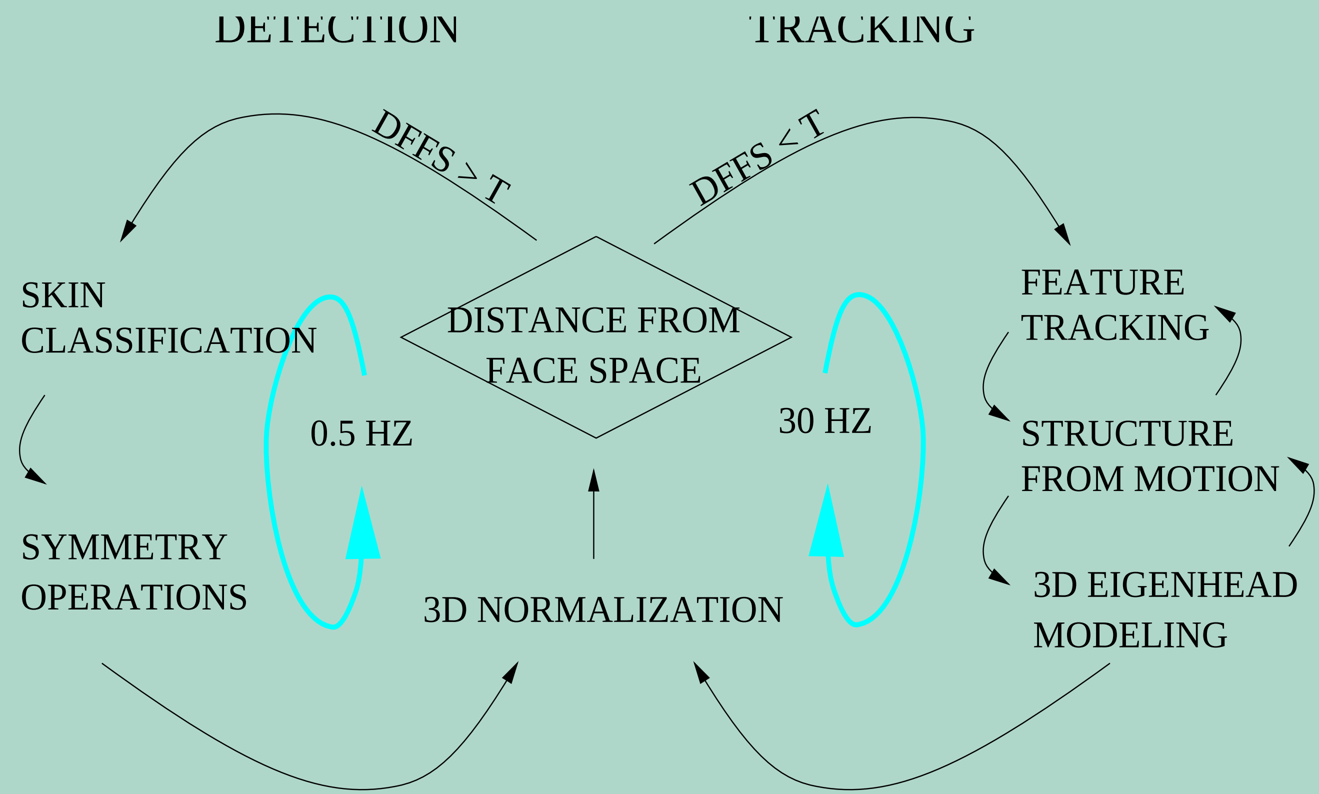


3D FACE TRACKING AND SfM

– In Real–Time, Detect, Track and Normalize a face in 3D by computing 3D pose and coarse 3D structure

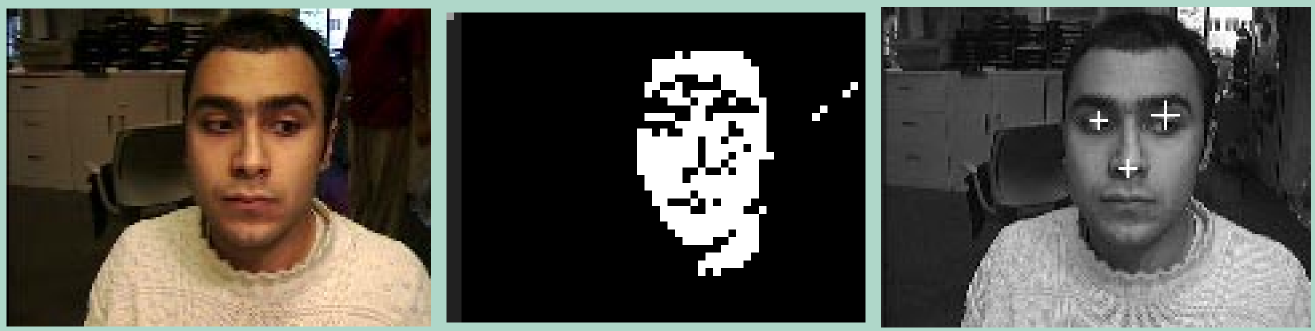
– 3D GLOBAL ADAPTIVE COUPLING OF 2D FEATURES

Local 2D Feature Tracking <–> Global Rigid 3D Structure Holistic Texture

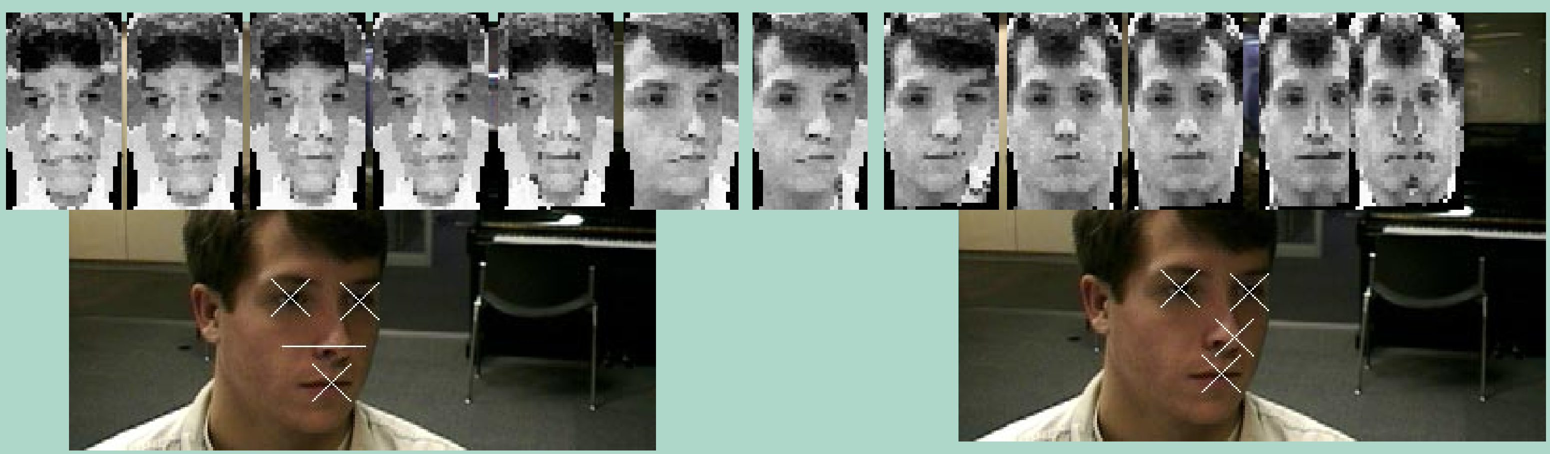


DETECTION AND DFFS

– YUV Mixture of Gaussians skin color model (EM)
– Symmetry Operations (Radial and Axial)
– Geometric Constraints



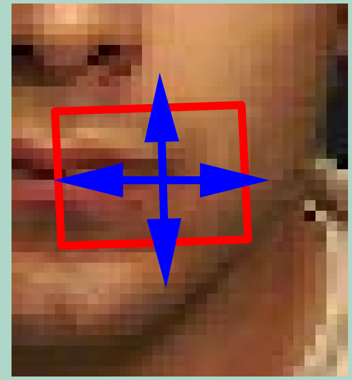
– Refining and Pruning initial localization with DFFS on 3D Frontally Warped Mug–Shot Images



2D FEATURE TRACKING

Initialize 8 Local 2D Feature Trackers

SSD NORMALIZED CORRELATION

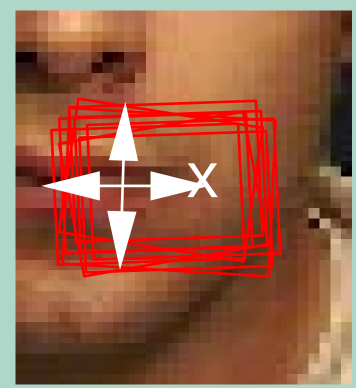


TRANSLATION

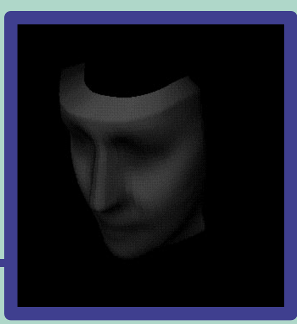
SCALE
ROTATION

–Get 2D motion, $u = (T_x, T_y, \theta, \text{scale})$ via linear model of SSD
–2D motion is also represented as motion of 2 anchor points

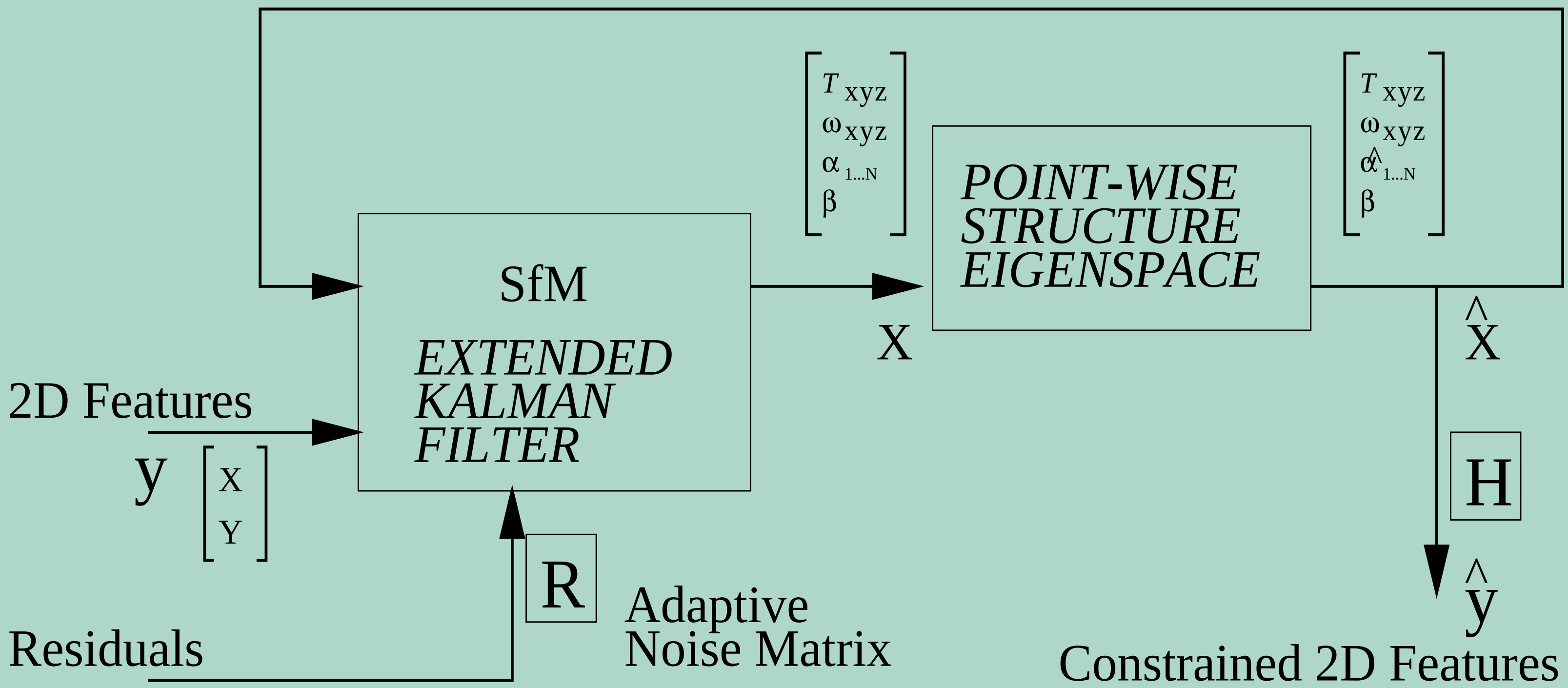
Also get sensitivity of model or 4x4 covariance of u (over SSD) by artificially perturbing each tracker.



$$C_{4x4}\ (u, SSD)$$



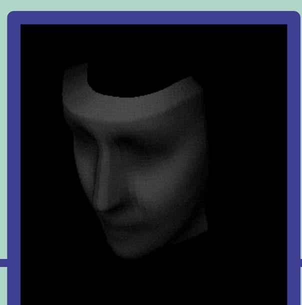
FEEDBACK STRUCTURE FROM MOTION



– Get 2D Features from anchor points on each patch
– Form block–diagonal $R = diag(C_1, C_2, \ldots C_8)$
– Form pointwise eigenspace of structure (α) from Cyberware

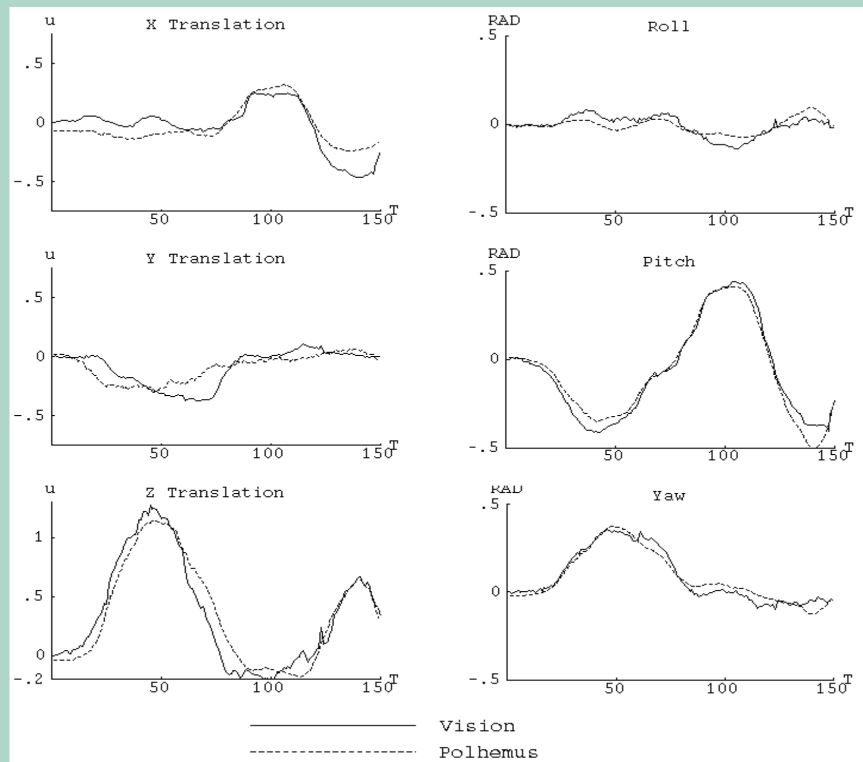
– Use feedback of constrained, rigid SfM to re–initialize the 2D trackers for the next frame.

– GLOBAL ADAPTIVE 3D FUSION OF 2D TRACKERS



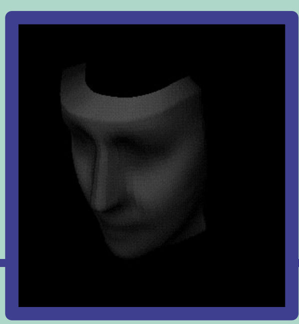
TRACKING ACCURACY & STABILITY

– Robust due to 3D SfM based integration of multiple trackers
– Recover 3D Translation, Rotation, Focal Length
Point–Wise Structure Estimate (20 points)
– 30 Hz Tracking (1 Hz detection)
– Stable in tracking mode for extended periods of time
– Robust to <65 degrees rot
large occlusion
fast motion



–Vision vs Polhemus Head Tracking and Struture Est.

RMS differences
0.11 units
2.35 degrees
Scale 10–12cm * 0.11

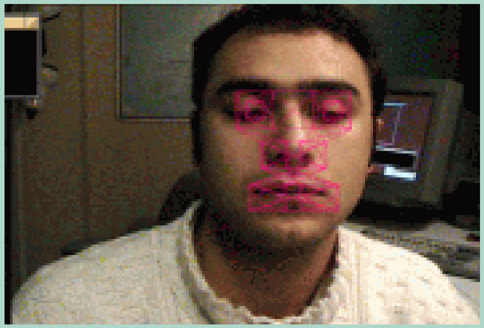


Within polhemus error [Azarbayejani & Pentland, 95]

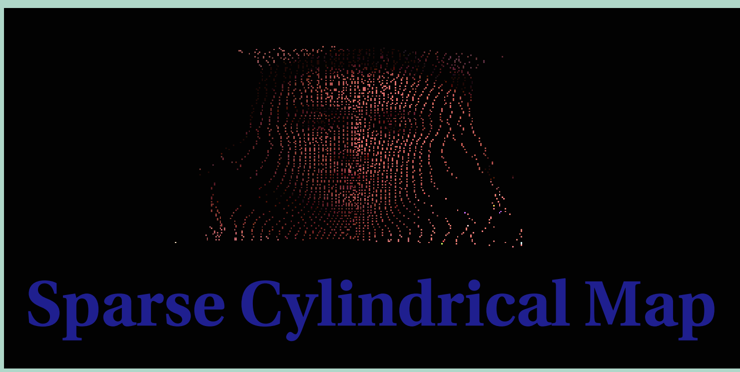
TEXTURE STABILIZATION & 3D REGRESSION

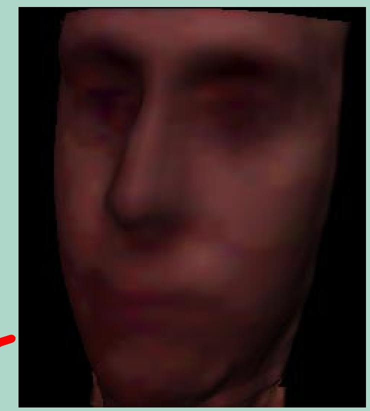
– Use estimates of Tx, Ty, Tz and quaternion to align a average Cyberware head to the face and texture map it

Tracking



Synthesized Frontal Mug–Shot

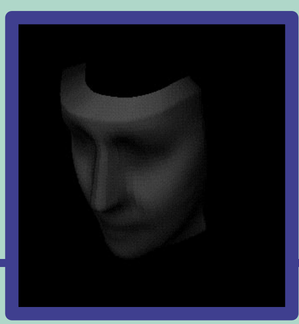




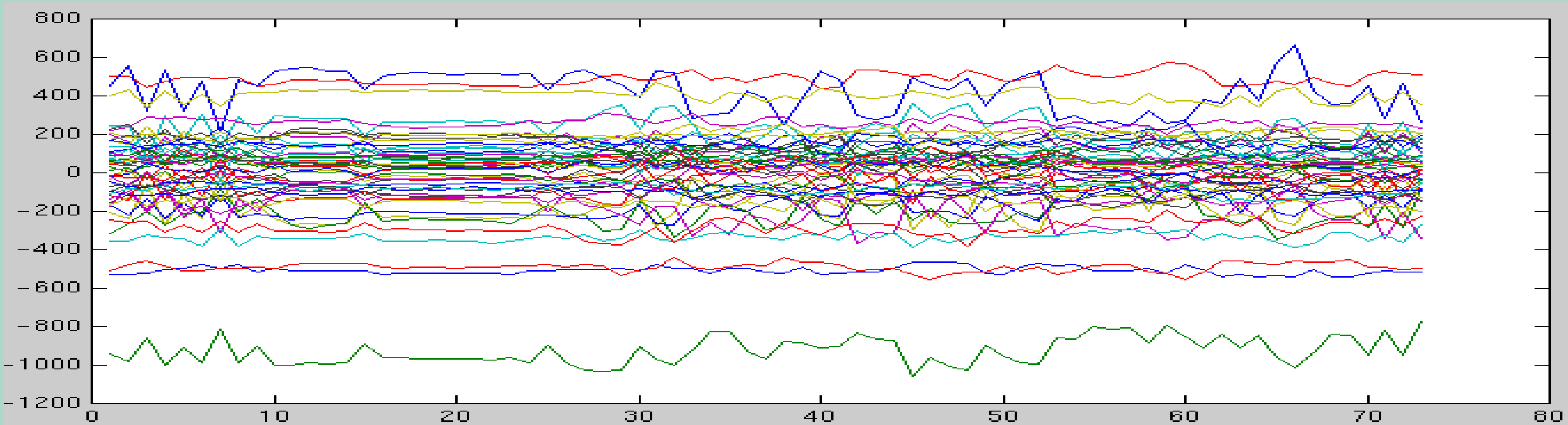
3D Model

Regress unknown components of sparse cylindrical map:

1) Continuous Texture (interpolate texture)
2) 3D Range Data Structure



TEMPORAL DESCRIPTION

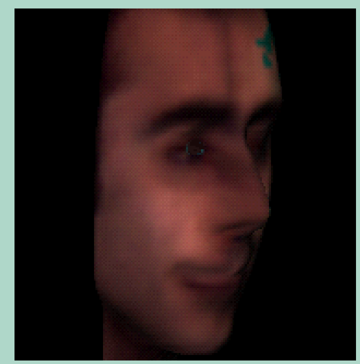


Coefficient Magnitudes

Time

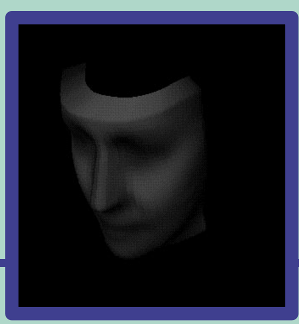
Static Face

Motion



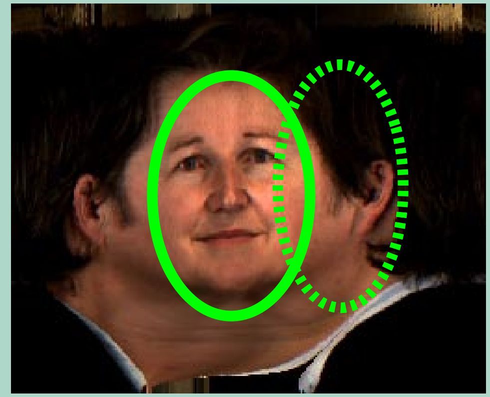
During Static Tracking:

Avg %SSD difference from Cyberware scan (in eigenspace) = 18 % Magnitude

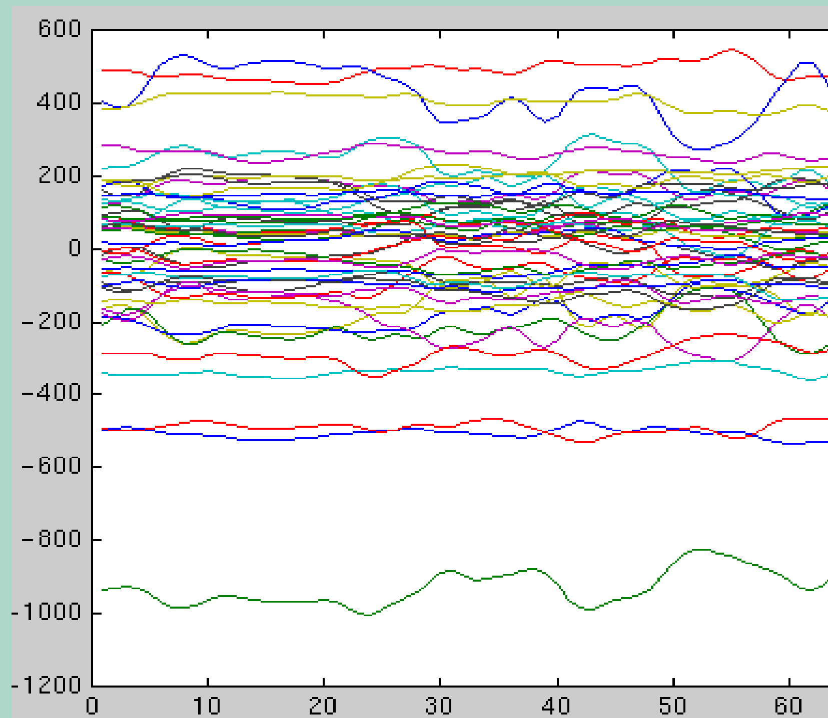


SPATIO–TEMPORAL INTEGRATION (KF)

CURRENT & FUTURE WORK:

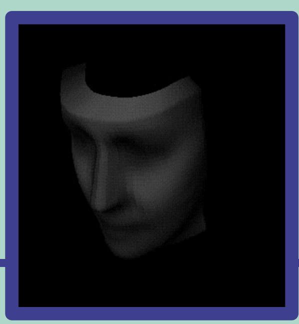


– Spatial Structure Integration
Accumulation of texture data over whole cylindrical structure when tracking large pose variations



–Tempral Structure Integration:
Recursive integration of eigenspace coefficients via Kalman Filter

– More flexible non–linear models



– More detailed temporal integration and intra–personal & extra–personal analysis