Head Tracking Using a Textured Polygonal Model

Arno Schödl, Antonio Haro, Irfan Essa
Computational Perception Lab
College of Computing, GVU Center
Georgia Institute of Technology

Motivation

- noninvasive tracking
- direction of attention
- head gesture recognition: nodding, shaking the head
- precursor to facial expression recognition

Previous Work

- Azarbayejani et. al. ‘93
- Black and Yacoob ‘95
- Basu, Essa and Pentland ‘96
- DeCarlo and Metaxas ‘96
- Jebara, Pentland ‘97
- La Cascia, Isidoro and Sclaroff ‘98

Key Properties

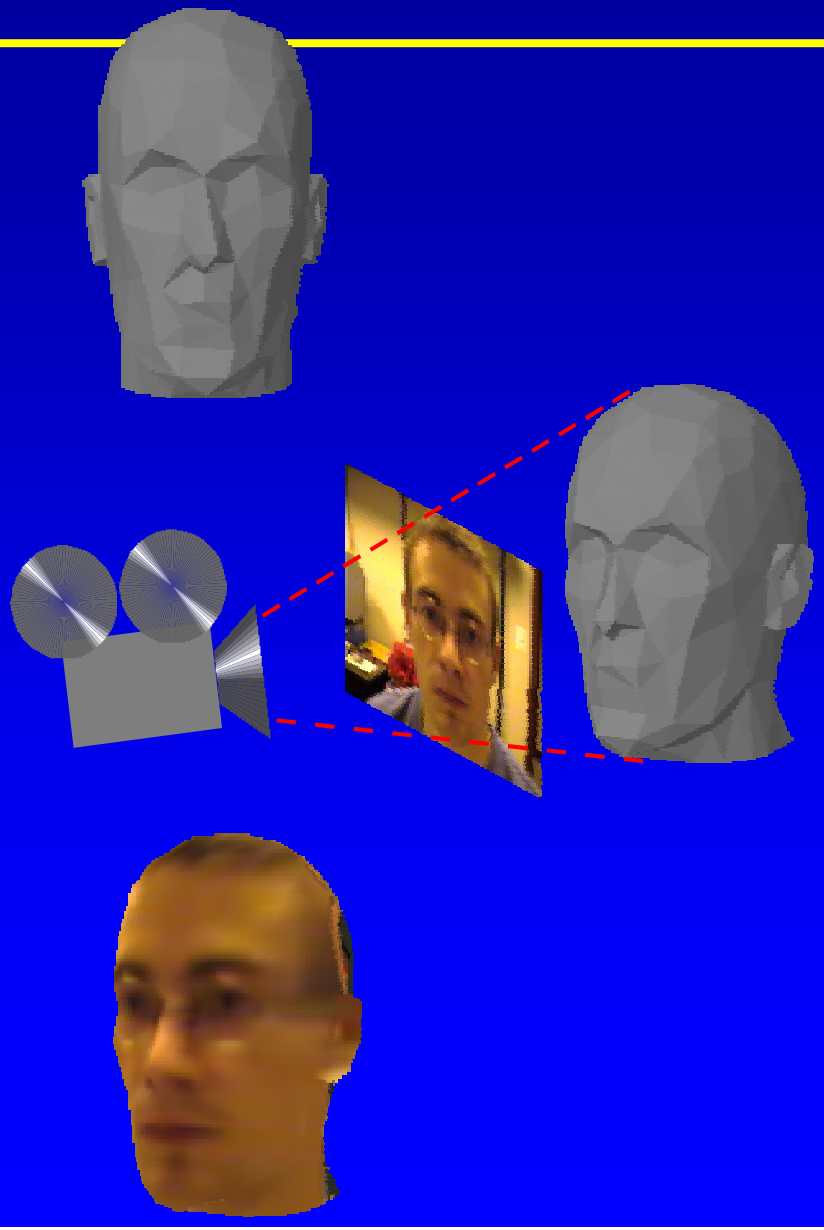
- no error accumulation
- robust against sudden large head movements
- general technique, applicable to any model-based rigid/nonrigid tracking

Initialization

① model

② alignment

③ texture



The Textured Model

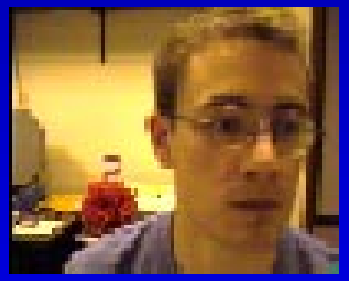
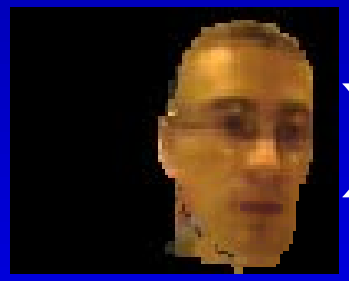
- only visible parts textured
- some geometric features as texture
- texture errors if the head shape is wrong

The Matching Process

- model image matched with camera
- iterative process, every step reduces difference
- when difference minimal, go to next camera frame

Our Error Function

- minimal when parameter values correct

$E = \rho($  -  $)$

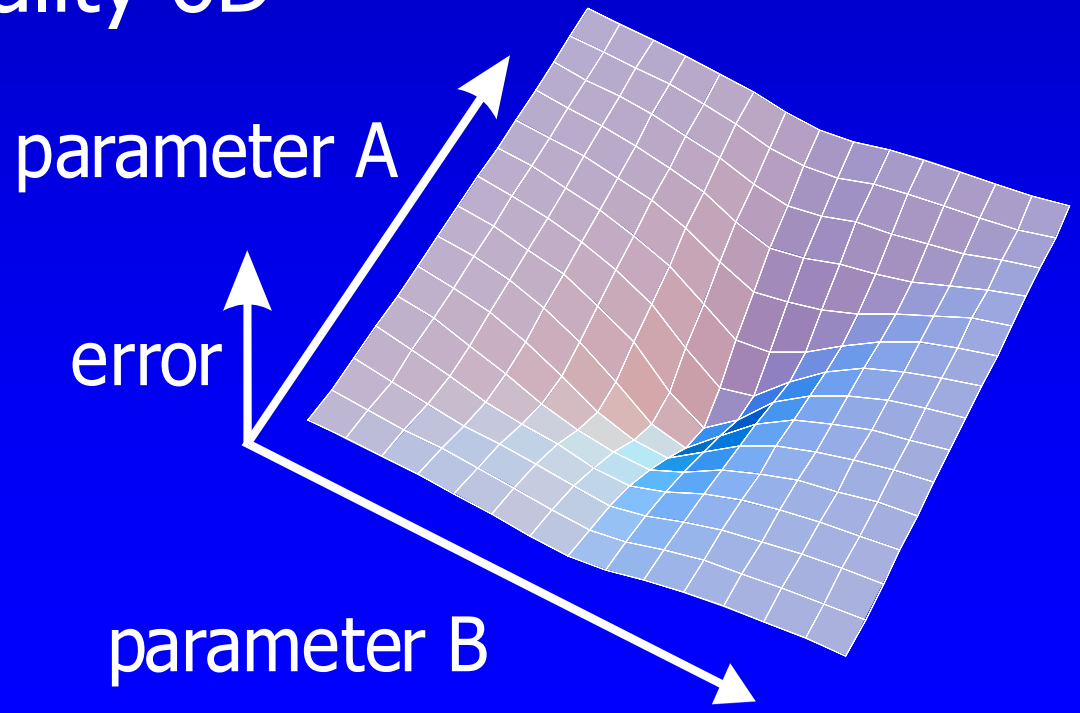
difference between
real image and model

robust norm

error

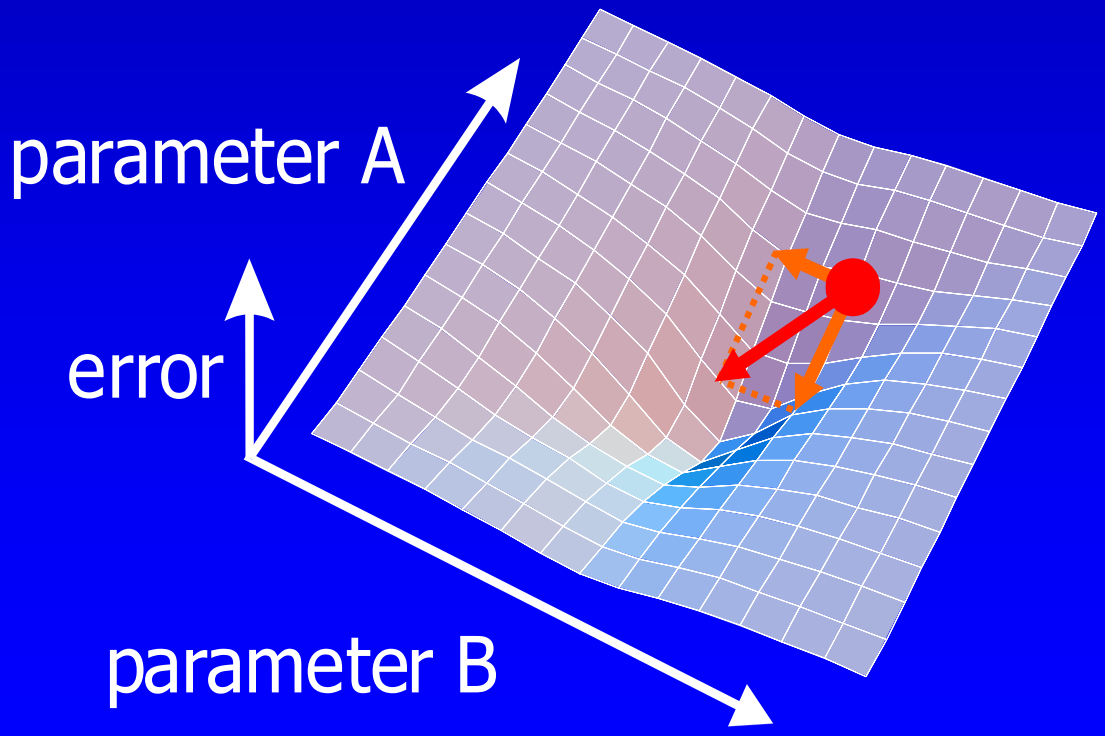
Error Minimization

- error function visualized as landscape
- find minimum by sliding downhill
- in reality 6D



The Gradient

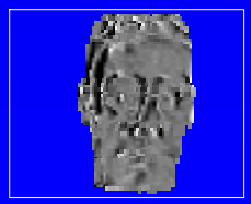
- minimization requires gradient
- gradient = (dE/dA, dE/dB)



Pixel Motion

- only look at pixels with texture
- calculate difference camera - model image

- calculate error change for pixel movement



horizontal

vertical

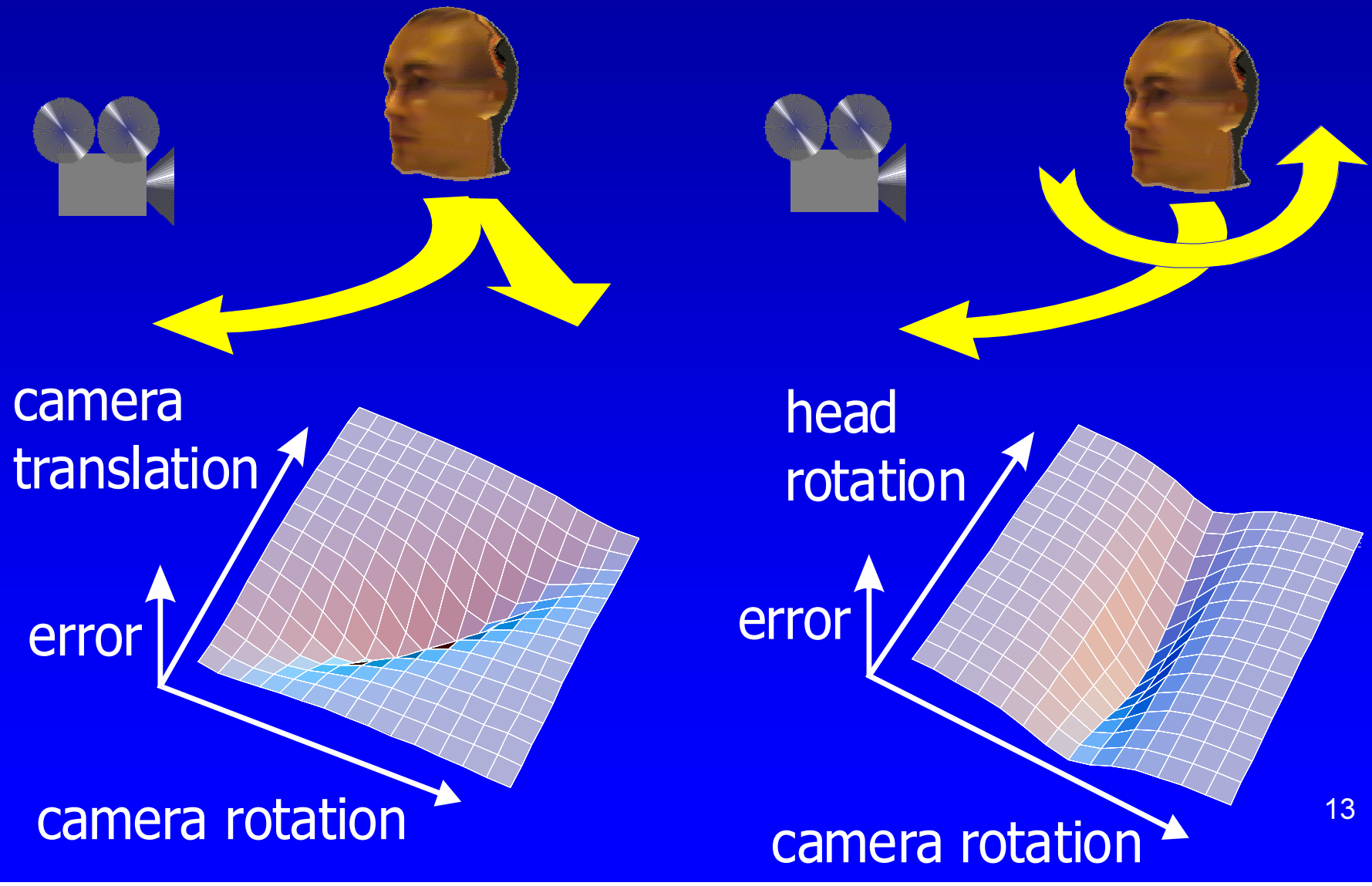
Gradient Computation

- example: rotation in screen plane
- motion pattern describes pointwise model motion
- multiplication with pixel motion yields gradient:

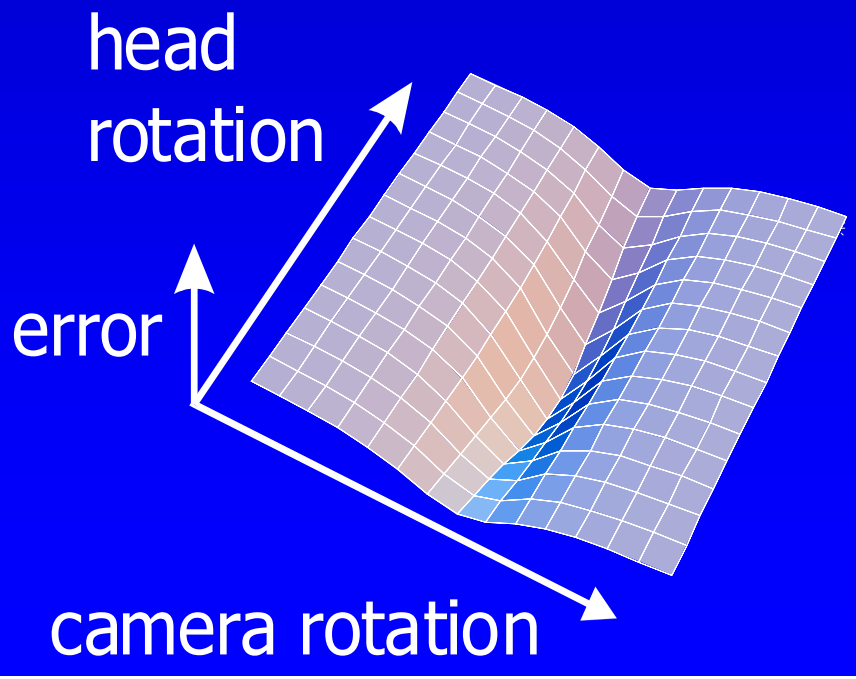
horizontal vertical

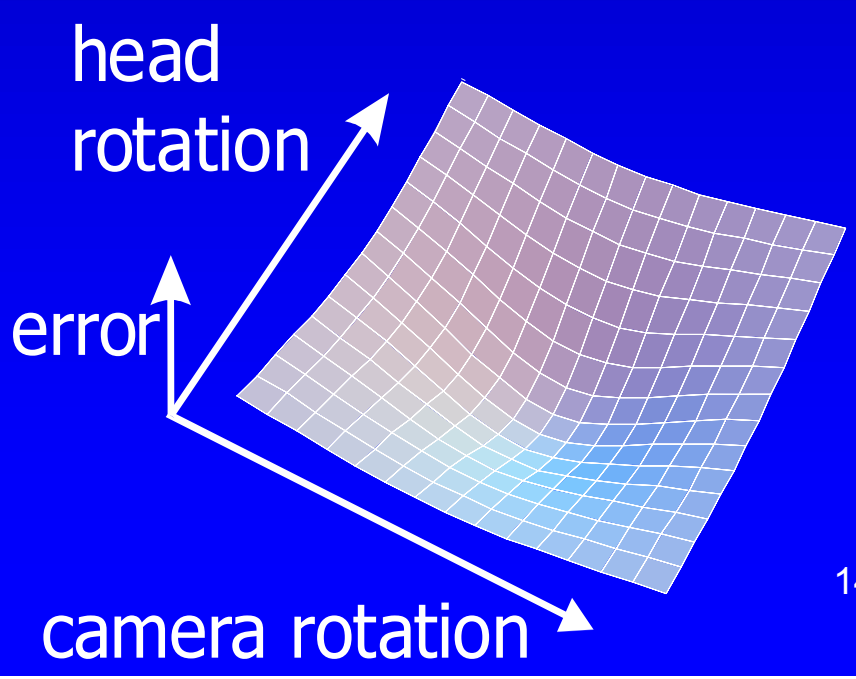
pixel motion | pattern | pixel motion | pattern | gradient

Parameterization



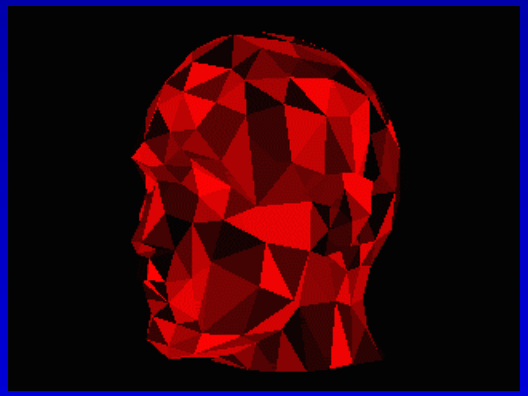
Parameterization (2)





Implementation

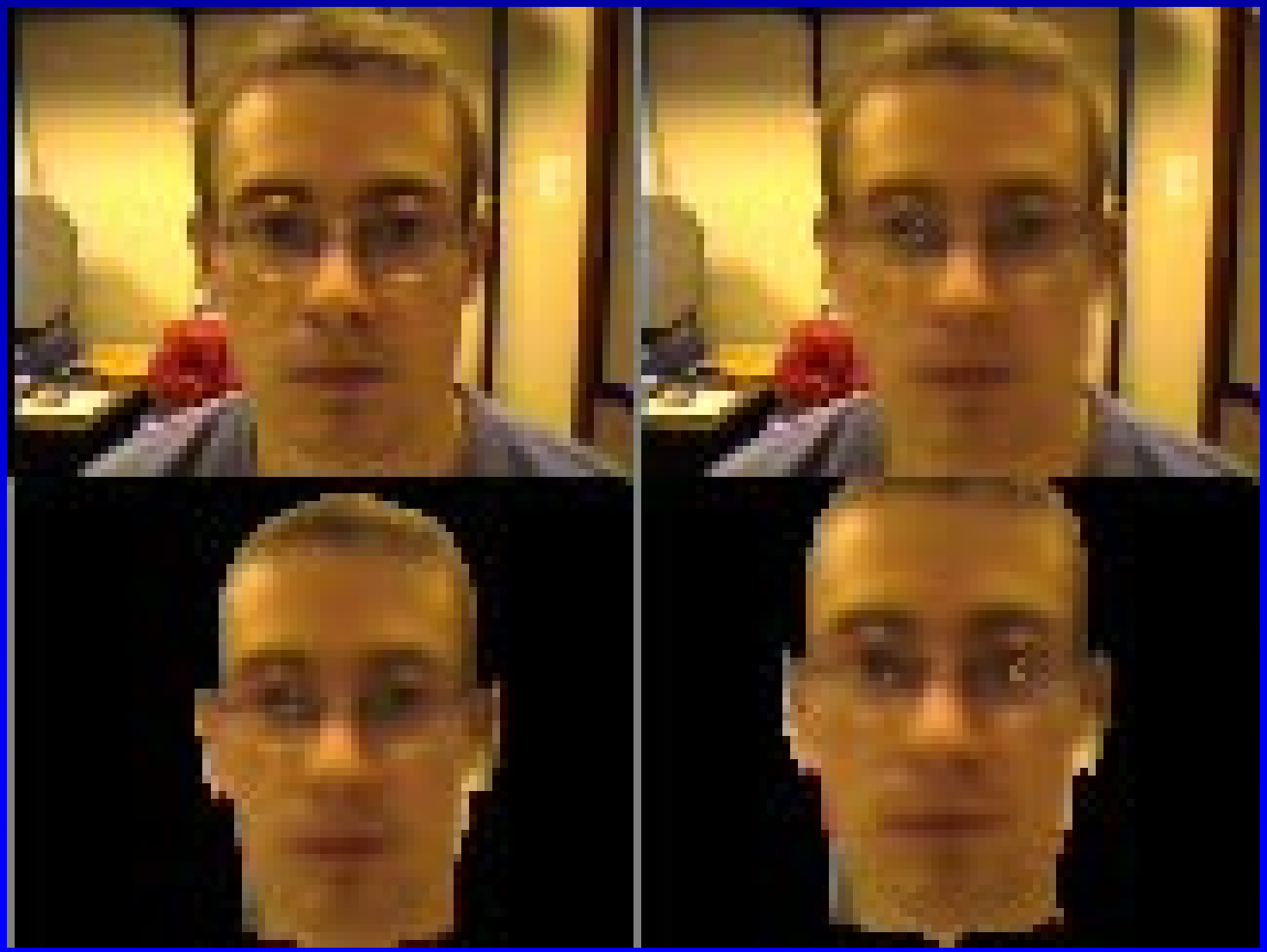
- Head model with 500 polygons
- OpenGL for model rendering and pixel-polygon-association
- 80 by 60 pixels resolution
- Conjugate Gradient with adaptive step size
- 2-stage Gaussian pyramid



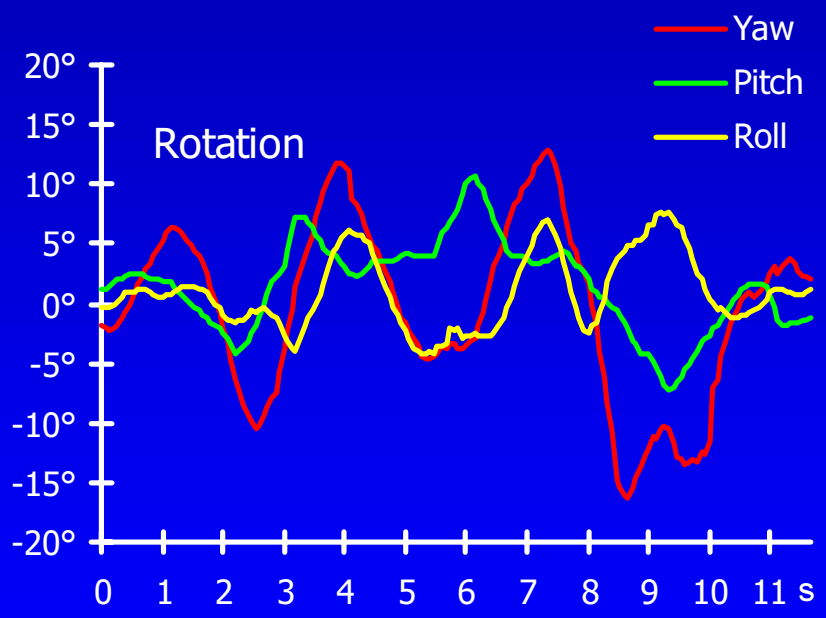
Results

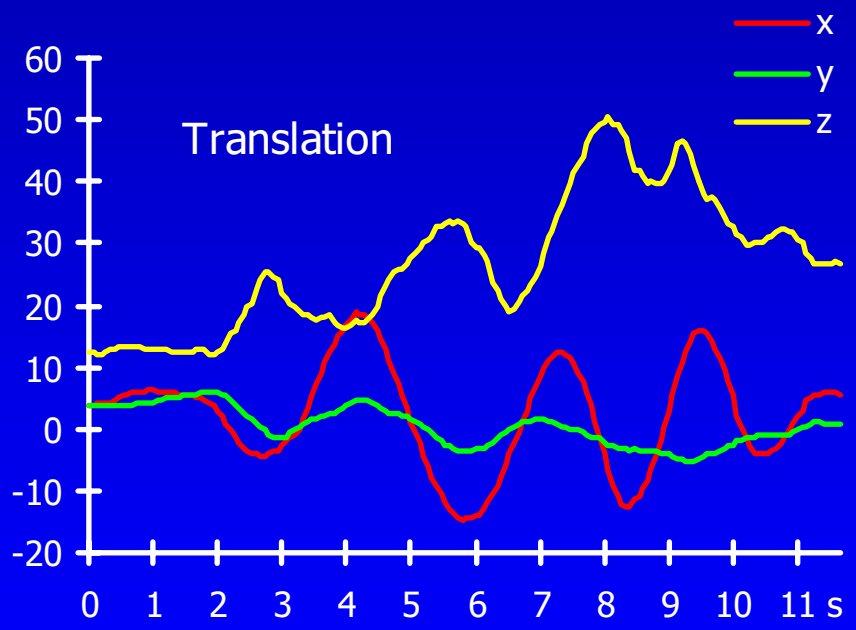
- Robustness
  - Rotation changes of 25 degrees
  - Translation changes of half the head size
- Speed
  - 1.7 s/frame on a Pentium II 300 Mhz
  - 0.25 s/frame on a 8 node PentiumPro 200 Mhz cluster (with K. Schwan)

Results



Results


Rotation
Yaw
Pitch
Roll
20°
15°
10°
5°
0°
-5°
-10°
-15°
-20°
0 1 2 3 4 5 6 7 8 9 10 11 s


Translation
x
y
z
60
50
40
30
20
10
0
-10
-20
0 1 2 3 4 5 6 7 8 9 10 11 s

Results

camera

model

overlaid model

stabilized

Limitations

- no texture for large head rotations
- strong facial expressions
- specular reflections and lighting changes

Future Work

- applying texture from multiple views
- adaptive head geometry
- applying model-based technique to other nonrigid tracking problems, e. g. facial expressions

Conclusion

- new technique for head tracking with a camera
- generic head model is textured with image of the user's head
- tracking by matching the rendering of the model with the real head in the video
- generally applicable to model-based tracking