Maximum-Likelihood Motion Estimation of a Human Face

Geovanni Martínez
University of Costa Rica
School of Electrical Egineering

Image Processing & Computer Vision Research Laboratory

IP&CV-LAB

http://ipcv-lab.eie.ucr.ac.cr

gmartin@eie.ucr.ac.cr

IEEE-ICME
July 2003

Introduction (1)

Challenge:

Non-intrusive human face motion estimation from a single video camera.

Applications:

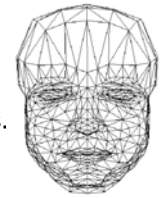
Video compression.
Human-machine interfaces.

Approach:

Motion estimation by maximizing the conditional probability of the frame to frame intensity differences at observation points.

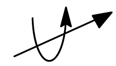
Introduction (2)

Shape model



The human face is described by a rigid 3D mesh of triangles.

Motion model



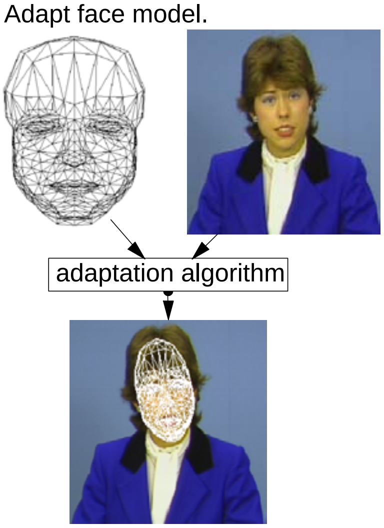
The global 3D motion of a human face is described by 6 parameters **B**: one 3D translation vector and 3 rotation angles.

Other assumptions

Static camera.
Perspective projection.
Diffuse illumination.

Motion estimation algorithm (1)

1) Read first image.
2) Adapt face model.



Motion estimation algorithm (2)

3) Create observation points.

Each observation point lies on the surface of the mesh and carries the corresponding intensity value at its position.

Only points with high linear intensity gradient are selected as observation points.

Motion estimation algorithm (3)

4) Read next image.

5) Compute the intensity differences **FD** between the observation points and the current image.

6) Compute the conditional probability p(**B**/**FD**).

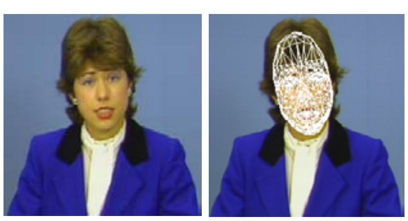
7) Maximize the conditional probability:

$$p(\mathbf{FD}/\hat{\mathbf{B}}) \geq p(\mathbf{FD}/\mathbf{B}) \;\; \forall \;\; \mathbf{B}.$$

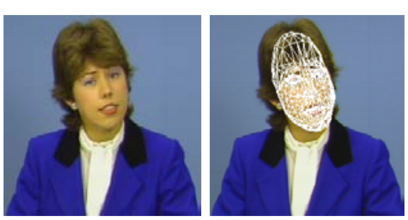
8) Move all vertices of the mesh and all observation points with the motion estimates $\hat{\mathbf{B}}$.

9) Goto step 4).

Results (1)

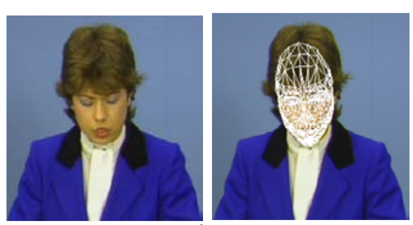


t_{13}

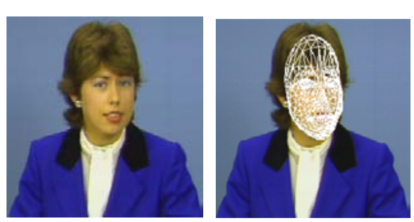


t_{30}

Results (2)



t_{71}



t_{149}

Conclusions

1) A non-intrusive human face motion estimation algorithm from a single video camera was developed.

2) The motion parameters of a human face are estimated by maximizing the conditional probability of the frame to frame intensity differences at observation points.

3) There are visible tracking errors particularly when the person opens or closes the eyes or mouth.