FACIAL ACTION TRACKING USING PARTICLE FILTERS AND ACTIVE APPEARANCE MODELS

Soumya Hamlaoui, Franck Davoine

HEUDIASYC - CNRS / Université de Technologie de Compiègne

Our objective

Tracking a near- frontal view face in video sequences:

- **Global motion:** 2D pose (position, scale, orientation)
- **Local motion:** facial features (appearance variations)

Proposed scheme

Stochastic tracking system based on the CONDENSATION algorithm (particle filtering):

- The unobserved state includes pose and appearance parameters.
- The observations distribution is derived from an Active Appearance Model (AAM) and uses a robust distance measure.
- The dynamic distribution and the particle number are adaptive.

UMR CNRS 6599 Heuristique et Diagnostic des Systèmes Complexes

State space & observations

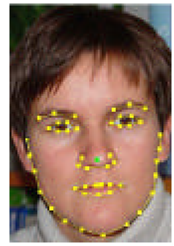
Unobserved state: $\boldsymbol{x}_t = \begin{pmatrix} pose_t \\ c_t \end{pmatrix}$

$pose_t = \boldsymbol{p}_t = (\mathrm{t}_x, \mathrm{t}_y, \delta, \theta)_t^{\mathrm{T}}$

c_t : first four modes of the appearance variation (92 % of appearance variations)

Observed data: $\boldsymbol{z}_t$ = image texture = $\mathbf{g}_{image}$

Active Appearance Models [Cootes & al 01]



PCA/ Shape: $\mathbf{s} = \mathbf{s}_m + f_s\ \mathbf{b}_s$

PCA/ Texture: $\mathbf{g} = \mathbf{g}_m + f_g\ \mathbf{b}_g$

Coupling the two models: $\mathbf{b} = \begin{pmatrix} W_s\, b_s \\ b_g \end{pmatrix}$

PCA / concatenated shape and texture parameters **b**: $\mathbf{b} = f_c\ \mathbf{c}$

Shape Instance: $\mathbf{s}_{model}(\mathbf{c}) = \mathbf{s}_m + \mathbf{Q}_s\ \mathbf{c}$

Texture Instance: $\mathbf{g}_{model}(\mathbf{c}) = \mathbf{g}_m + \mathbf{Q}_g\ \mathbf{c}$

Match a target face in a given image (iterative gradient search):

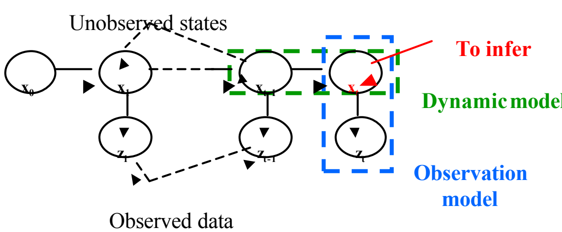
Minimize a texture residual: $r(\boldsymbol{c}, \boldsymbol{p}) = g_{model}(\boldsymbol{c}) - g_{image}(\boldsymbol{c}, \boldsymbol{p})$

Find the optimal correction ($\delta\boldsymbol{c}$, $d\boldsymbol{p}$) to apply in order to minimize $r(\boldsymbol{c}, \boldsymbol{p})$

$\delta\boldsymbol{c} = -R_c\, r(\boldsymbol{c}, \boldsymbol{p})$ $\quad\delta\boldsymbol{c} = -R_p\, r(\boldsymbol{c}, \boldsymbol{p})$

R_p , R_p Matrices precomputed from training data

(Condensation algorithm [Isard & Blake 98])

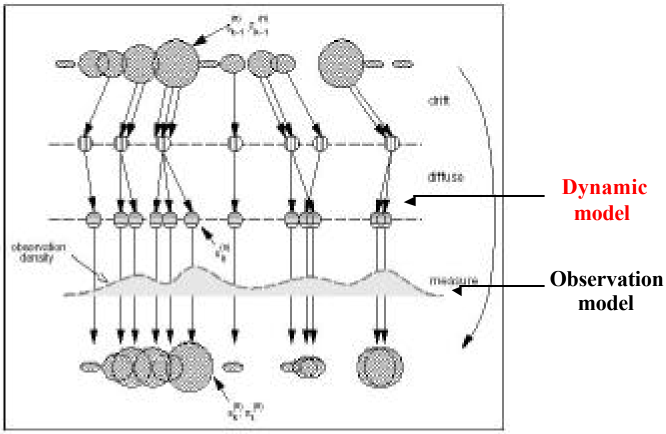


Tracking ° recursive evaluation of the posterior density of the target state conditionally to the history of observations $P(\mathbf{x}_t \mid \mathbf{z}_{1:t})$

by means of the empirical distribution of a system of particles :

- **Dynamic model**: $P(\mathbf{x}_t \mid \mathbf{x}_{t-1})$
- **Observation model**: $P(\mathbf{z}_t \mid \mathbf{x}_t)$

Description of the Condensation algorithm



Dynamic model (1/2)

Propagates the particles system through time :

$$x_t = \hat{x}_{t-1} + v_t + S_t u$$

$\hat{\boldsymbol{x}}_{t-1}$: estimate of the state vector at the previous time step.

$\boldsymbol{v}_t = (\partial \mathbf{p}, \partial \mathbf{c})^T$: predicted shift in pose/appearance obtained by an AAM search in the current frame.

$\boldsymbol{u}$: random variates having zero mean and unit variance.

$\boldsymbol{S}_t = \mathrm{diag}\,(s_t^{(t_x)}, \ldots, s_t^{(c_4)})$: standard deviations for each pose/appearance parameter.

Dynamic model (2/2)

- **Adaptive standard deviations** [Zhou & al 04]:

$$[s_t^{(t_x)},\dots,s_t^{(c_4)}]^T = diag(R_t^{(t_x)},\dots,R_t^{(c_4)})[s_0^{(t_x)},\dots,s_0^{(c_4)}]^T$$

$$R_t^{(i)} \equiv \sqrt{e}$$

$\sqrt{e}$: texture error averaged over the L pixels of the textures:

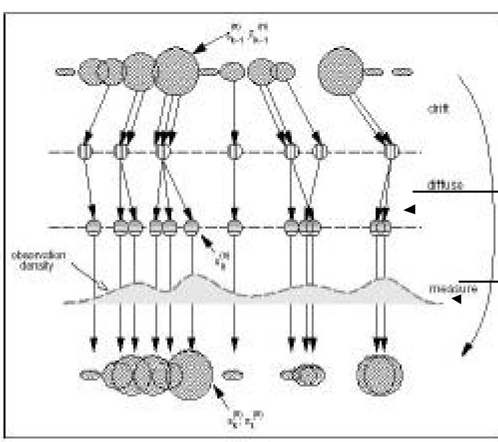
$$e_t = \frac{2}{L}\sum_{l=1}^{L} r\left(\frac{g^l_{model} - g^l_{image}(\tilde{x}_t)}{s_l}\right)$$

- **Adaptive particle number** (Substantial gain in computing time):

$$N_t = N_0 \sum_{i=1}^{8} R_t^{(i)}$$

N_0: initial fixed particle number

Description of the Condensation algorithm



Dynamic model

Observation model

Observation model (1/2)

Consists of the likelihood $P(z_t \mid x_t)$ according to which the particles are weighted:

Based on the difference between:

- **Image texture** : $g_{image}(p_t, c_t)$ sampled at the hypothesized pose and shape
- **Model texture** : $\mathbf{g}_{model}(c_t)$ given by the AAM

$$P(z_t \mid x_t) = C\,e^{-d[g_{model};g_{image}]}$$

C: Normalizing constant of this distribution

Observation model (2/2)

The texture distance *d(,)* is an error measure summed over all *L* pixels of both textures:

$$d(g,g') = \sum_{l=1}^{L} r\left(\frac{g_l - g_l'}{s_l}\right)$$

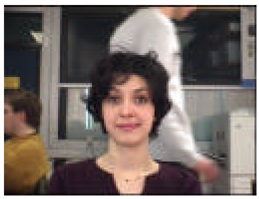
r() is a robust error function in order to reduce the influence of occluded pixels:

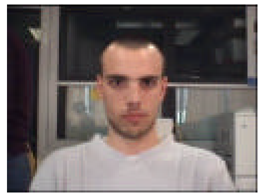
$$r(g) = \begin{cases} \frac{1}{2}g^2 & \text{if } |g| \le h \\ h|g| - \frac{1}{2}h^2 & \text{if } |g| > h \end{cases}$$

h: fixed threshold above which the difference $|g|$ is considered to be an outlier

Experimental results

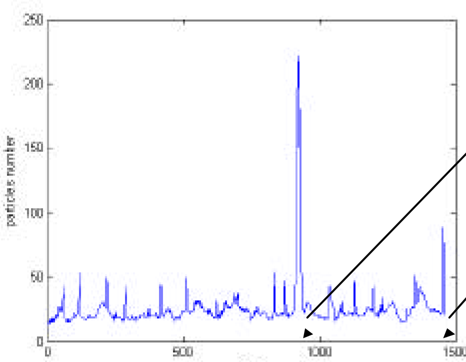
The proposed method was implemented in C++ and tested on a PC running WinXP at 2.4 GHz with 512 Mb of RAM.

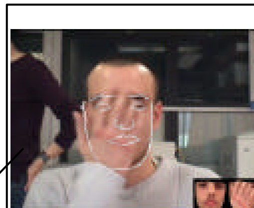




- $N_0 = 500$
- N_t evolves between about 20 and 80 and increase when change in pose and/or appearance is rapid
- 2 frames per second

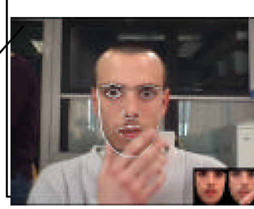
Particle number evolution





Frame 920

Nearly full occlusion



Frame 1445

Partial occlusion

But…

• The effectiveness of the appearance model remains conditioned by the fact that the tracked appearance must be beforehand learned and modelled.

• This modelling is sensitive to the recording conditions of the training images.

Replace the appearance model by an adaptive appearance estimated on-line

New texture model

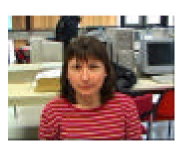
- $g_{on\text{-}line\ model}$: initialized manually using the face texture in the first frame

- Updated: $g_{on\text{-}line\ model}(t) = a\, g_{on\text{-}line\ model}(t-1) + (1-a)\, g_{image}(t, \tilde{x}_{t-1})$

$$\begin{cases} \alpha\text{: forgetting factor determining the update importance} \\ g_{image}(t, \tilde{x}_{t-1})\text{: the current image texture estimated to the state hypothesis at } t\text{-}1 \end{cases}$$

- The hidden state space encodes the pose and the first four modes of the shape parameters obtained from the face model:

$$x_t = (p_t, b_s)^T$$



Some perspectives

- Recognition of facial actions and behavior, by analysing the state trajectories (applications: HCI, surveillance, etc.).
- 3D pose tracking.

Thank you for your attention

[Cootes & al 01] : T.F. Cootes, G.J. Edwards and C.J. Taylor. *Active Appearance Models*. IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 681-685, June 2001.

[Hamlaoui & Davoine 05] : S. Hamlaoui, F. Davoine. *Facial action tracking using an AAM-based Condensation approach.* IEEE Int. Conf. on Acoustic, speech and Signal Processing, March 2005.

[Isard & Blake 98] : M. Isard, A. Blake. *Condensation – Conditional Density Propagation for Visual Tracking.* Int. Journal of Computer Vision, pp. 5-28, 1998.

[Zhou & al 04] : S. Zhou, R. Chellappa, and B. Moghaddam. *Visual tracking and recognition using appearance-adaptive models in particle filters*. IEEE Trans. on Image Processing, November 2004.