GFlow: A Generative Model for Fast Tracking using 3D Deformable Models.

(Cite as MPLab TR 2003.01 v3)

http://mplab.ucsd.edu

Javier R. Movellan, John Hershey

Tim Marks & Cooper Roddey

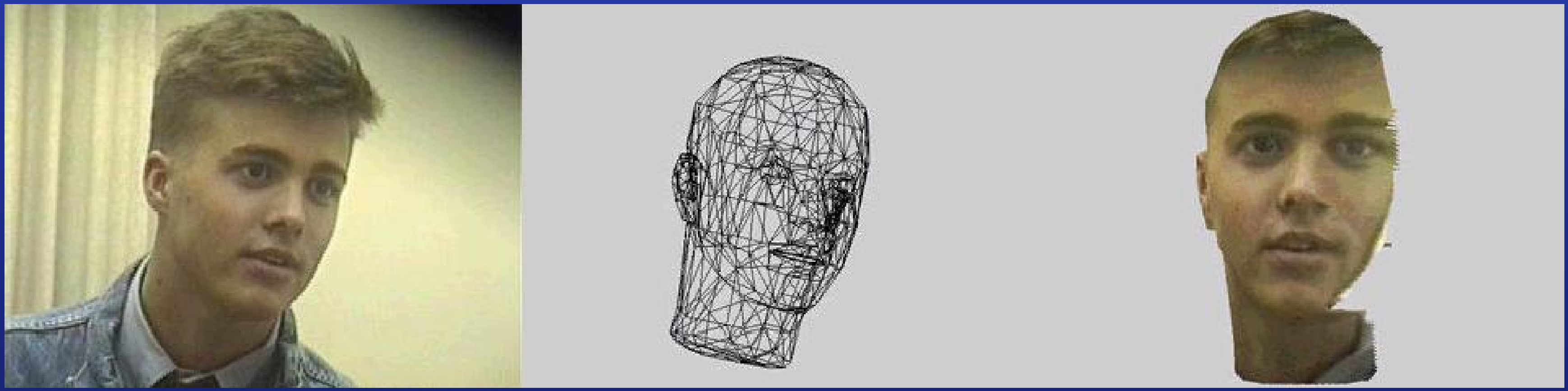
Machine Perception Laboratory

University of California San Diego

Dealing with Pose

- Multiple cameras.
- 3D Morphable models.
- Ensemble of pose specific detectors.

3D Morphable models



Bartlett, Braathen, Littlewort, Smith, Movellan (2001)

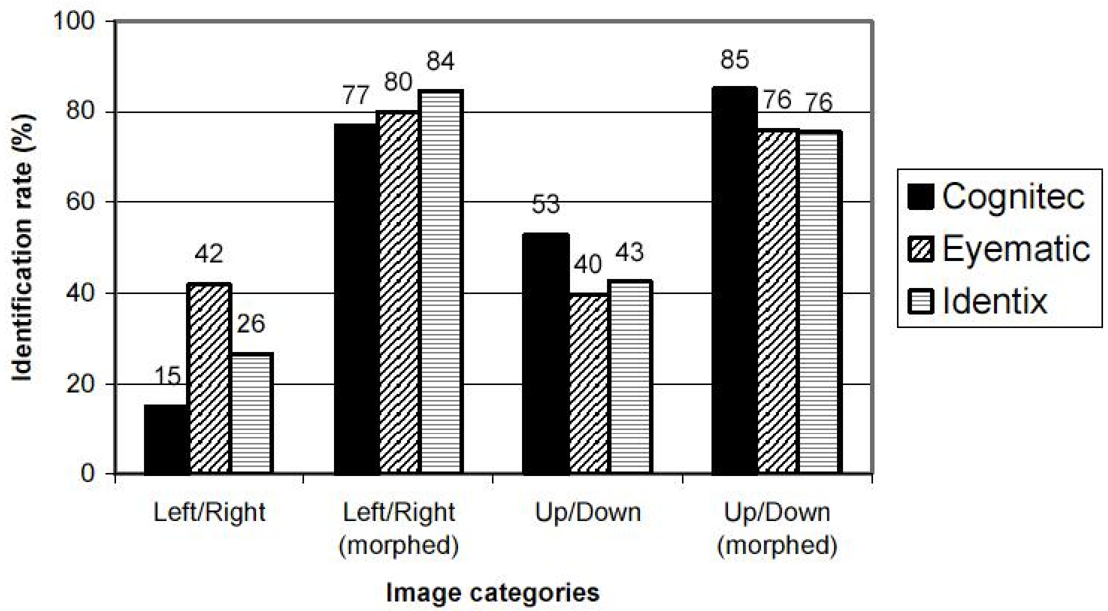


Figure 9. Identification performance is shown on non-frontal and morphed non-frontal images. The left/right and up/down categories are top identification rates for the original non-frontal images. The left/right (morphed) and up/down (morphed) categories are top identification rates for the morphed non-frontal images. Performance is on a database of 87 individuals.

Results FRVT02

Results FRVT02

3D Tracking: Current Approaches

- Optic Flow Approaches: Given two images y_t and y_{t+1} and the position of the object at time t estimate the position of the object at time $t+1$.
  - ⋆ Few assumptions about appearance of object.
  - ⋆ Good knowledge about location of object. Tendency to drift.

- **Template Based Approaches:** Given a template of the object appearance find it on the image plane.
  - ⋆ Few assumptions about location of object.
  - ⋆ Good knowledge of object appearance: Difficult to handle realistic sources of variation.

- **Template Based Approaches:** Given a template of the object appearance find it on the image plane.
  - ⋆ Few assumptions about location of object.
  - ⋆ Good knowledge of object appearance: Difficult to handle realistic sources of variation.

In practice people use heuristic combinations of template and flow:

Brand & Bhotika (2001)

Torresani, Yang, Alexander & Bregler (2001)

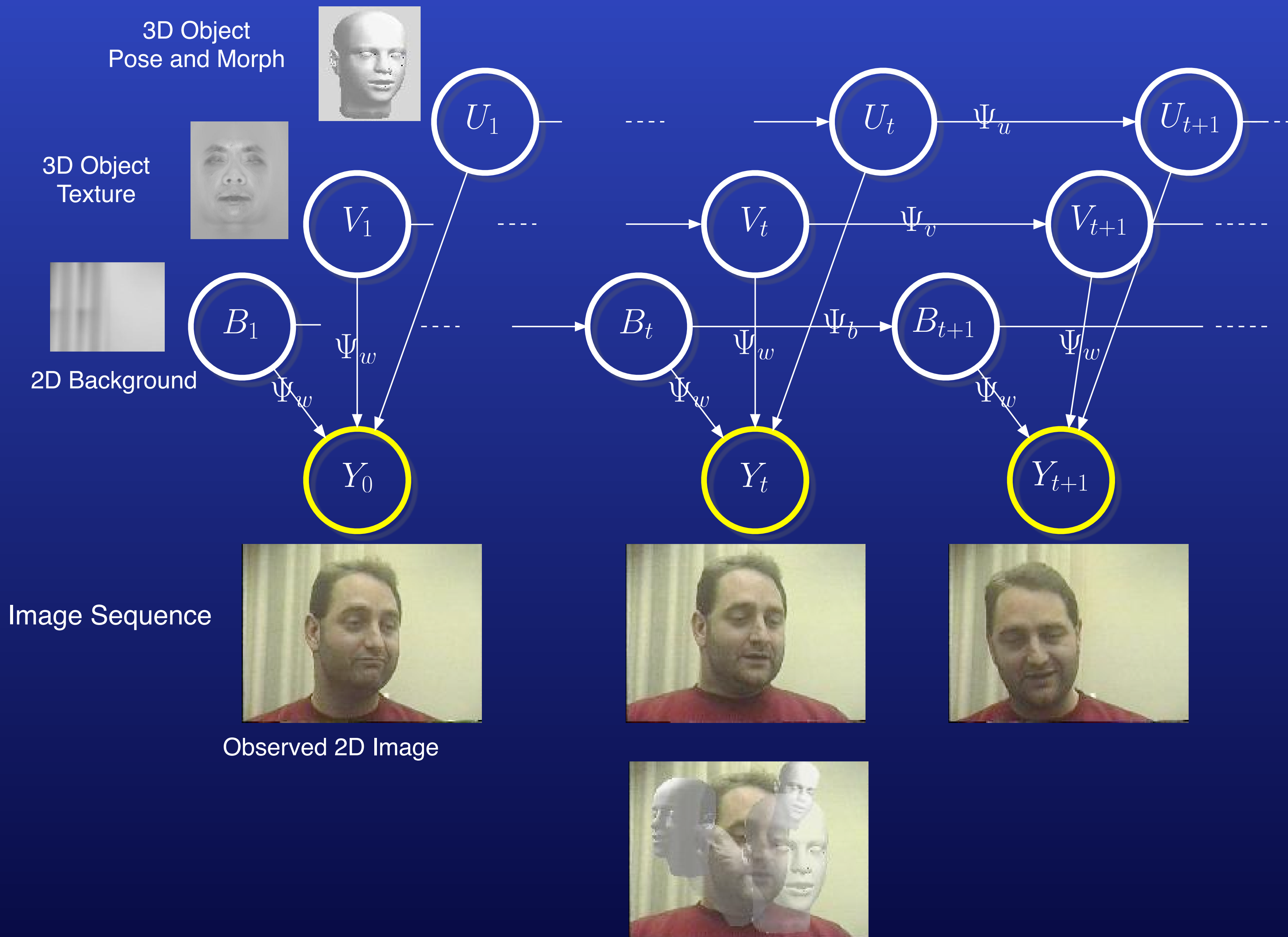
La Cascia & Sclaroff (2000)

Xiao, Kanade & Kohn (2002)

GFlow:

A generative model for tracking morphable objects.

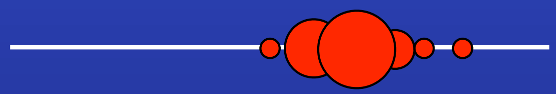
- Principled (Optimal Inference).
- Fast.
- Template and flow based approaches emerge as special cases.
- Uses foreground and background information.
- Easy to connect to other generative models (e.g. ICA.).



We want $P(u_t v_t b_t \,|\, y_1 \cdots y_t)$

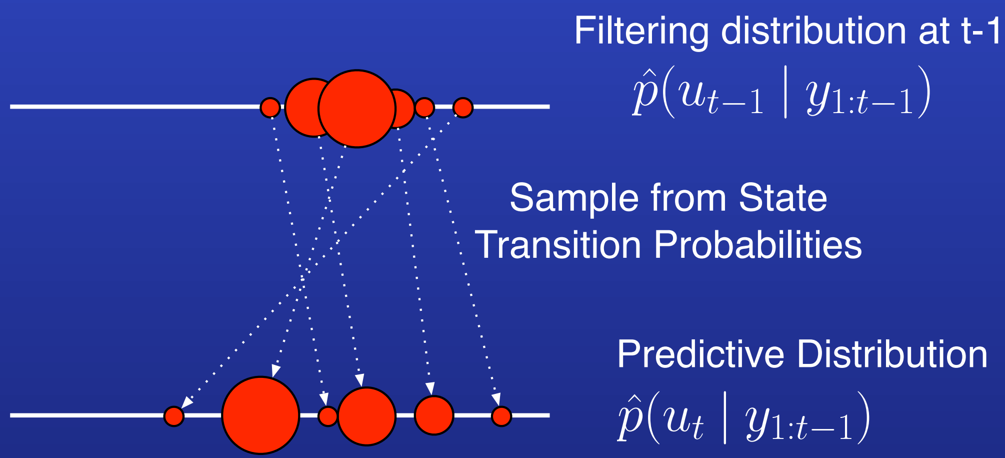
Non-Linear Filtering Problem

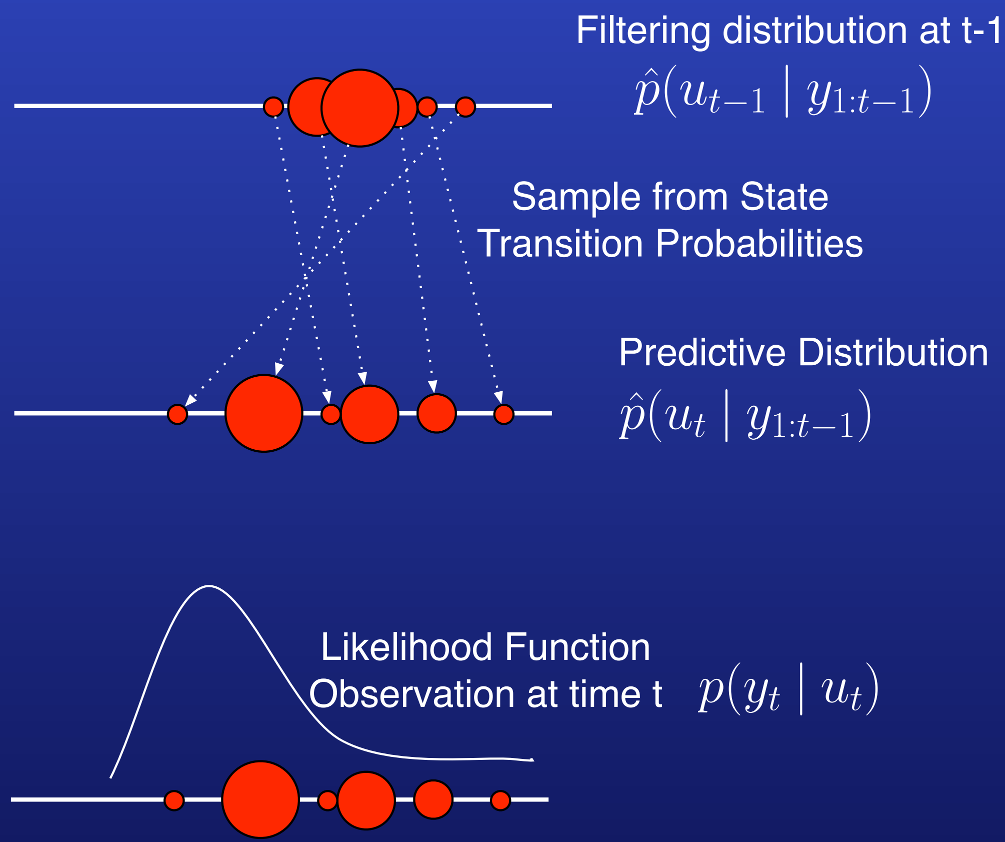
- Extended Kalman Filter (unimodal).
- Stochastic Partial Differential Equations.
- Discretizing hypothesis space (see dumbicles).
- Sampling (Particle Filters).

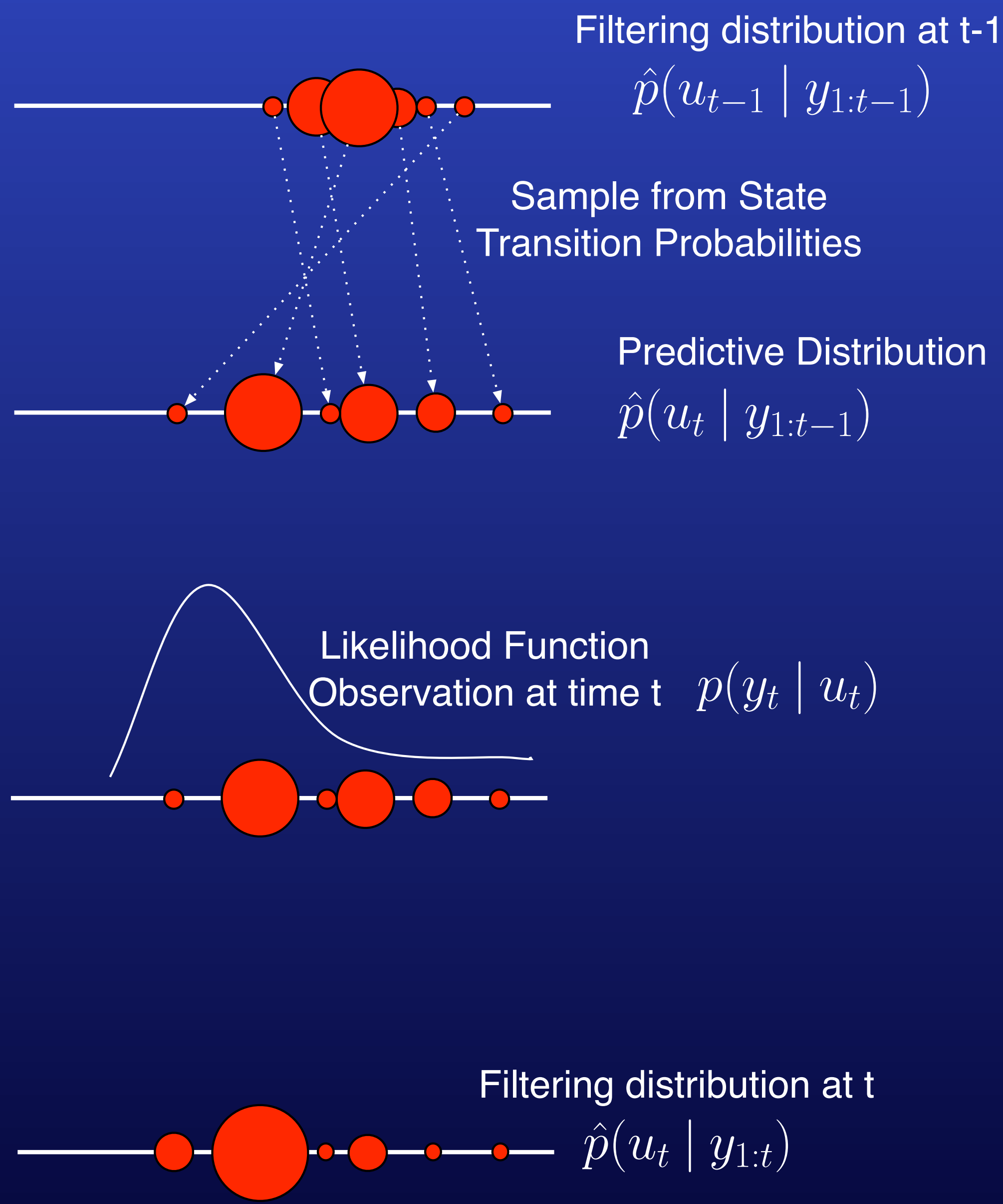


Filtering distribution at t-1

$$\hat{p}(u_{t-1} \mid y_{1:t-1})$$

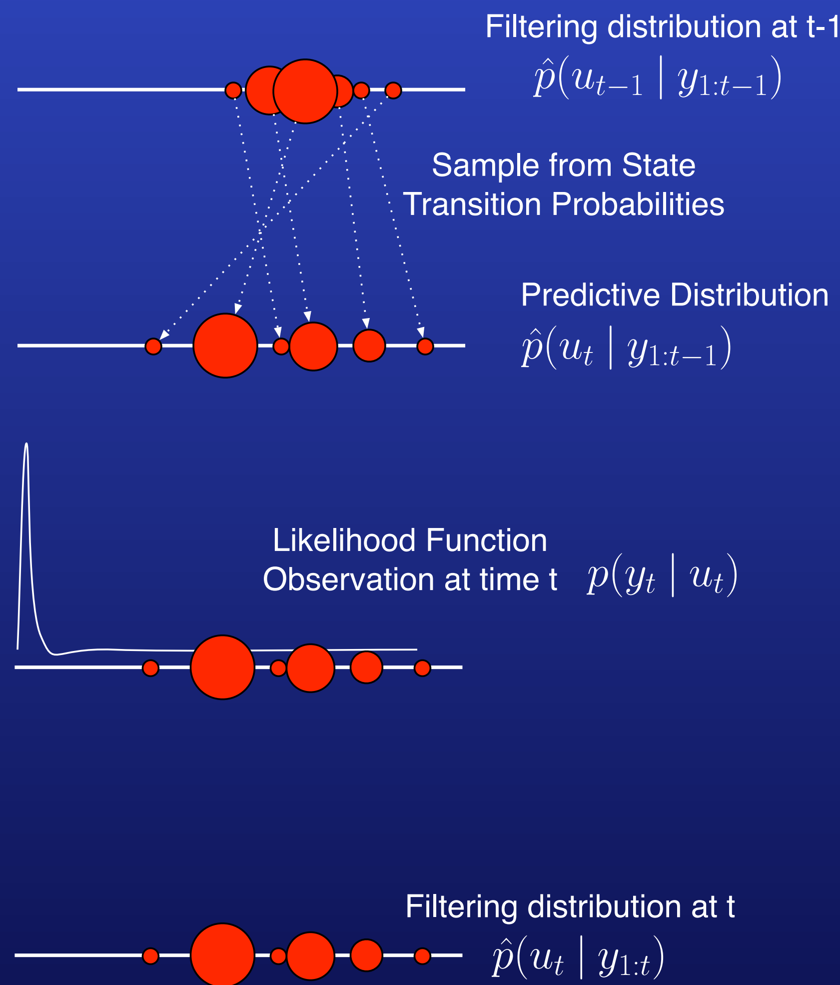

Filtering distribution at t-1
$\hat{p}(u_{t-1} \mid y_{1:t-1})$
Sample from State
Transition Probabilities
Predictive Distribution
$\hat{p}(u_t \mid y_{1:t-1})$


Filtering distribution at t-1
$\hat{p}(u_{t-1} \mid y_{1:t-1})$
Sample from State
Transition Probabilities
Predictive Distribution
$\hat{p}(u_t \mid y_{1:t-1})$
Likelihood Function
Observation at time t
$p(y_t \mid u_t)$

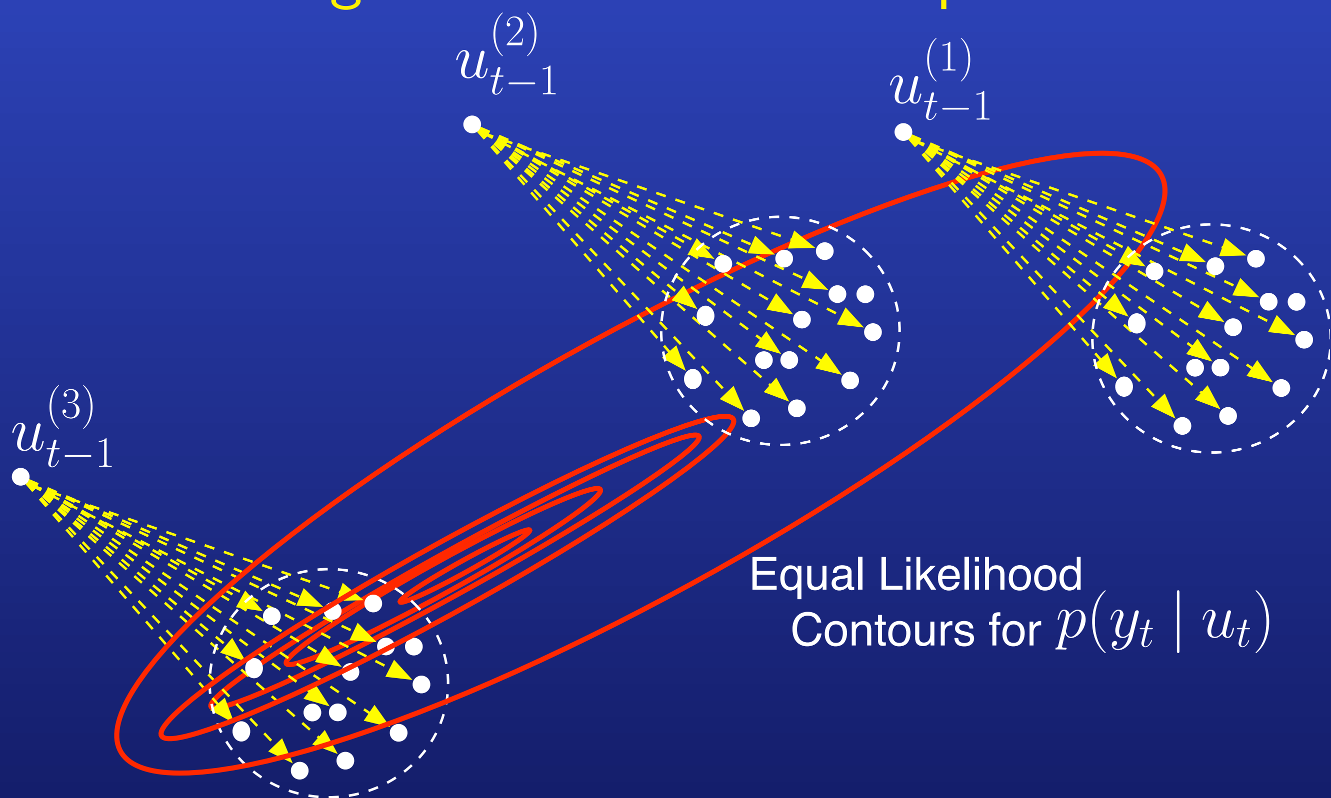


Filtering distribution at t-1
$\hat{p}(u_{t-1} \mid y_{1:t-1})$
Sample from State
Transition Probabilities
Predictive Distribution
$\hat{p}(u_t \mid y_{1:t-1})$
Likelihood Function
Observation at time t
$p(y_t \mid u_t)$
Filtering distribution at t
$\hat{p}(u_t \mid y_{1:t})$

The Needle in a Haystack Problem


Filtering distribution at t-1
$\hat{p}(u_{t-1} \mid y_{1:t-1})$
Sample from State
Transition Probabilities
Predictive Distribution
$\hat{p}(u_t \mid y_{1:t-1})$
Likelihood Function
Observation at time t $p(y_t \mid u_t)$
Filtering distribution at t
$\hat{p}(u_t \mid y_{1:t})$

Problem gets worse with more parameters



Conditionally Gaussian Problem

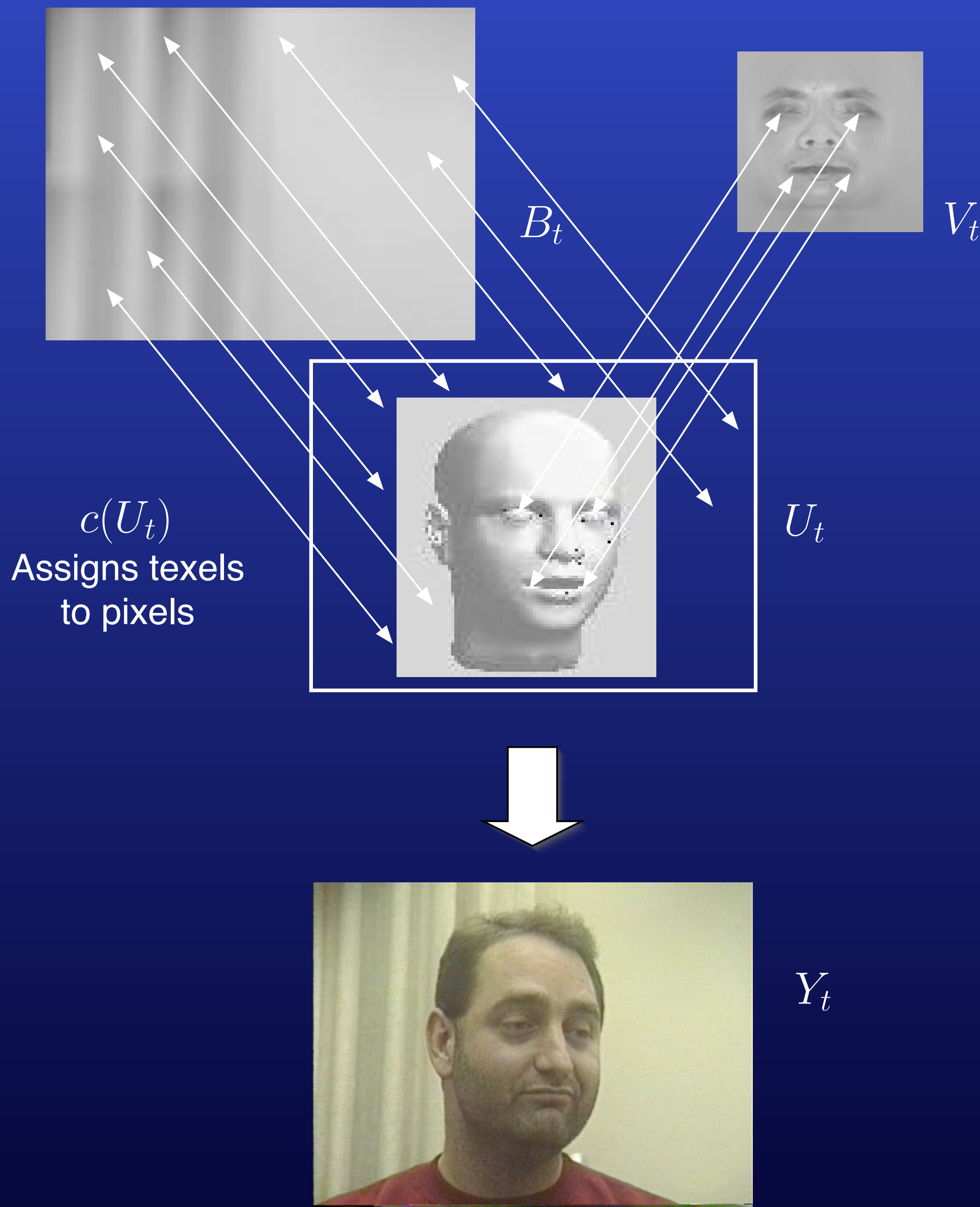
$$U_t \sim p(U_t \,|\, U_{t-1})$$ 3D pose and expression

$$V_t = V_{t-1} + Z_t^v$$ Object texture

$$B_t = B_{t-1} + Z_t^b$$ Background texture

$$Y_t = c(U_t) \begin{pmatrix} V_t \\ B_t \end{pmatrix} + W_t$$ Image

If we knew $u_{1:t}$ problem would be linear.


B_t
V_t
$c(U_t)$
Assigns texels
to pixels
U_t
Y_t

Expert Filtering

Sum Expert Credibility $\times$ Expert Opinion

Opinion of expert centered at u_{t-1}

$$p(u_t v_t b_t \mid y_{1:t}) = \int p(u_t v_t b_t \mid u_{1:t-1} y_{1:t}) p(u_{1:t-1} \mid y_{1:t}) du_{1:t-1}$$

Credibility of expert centered at u_{t-1}

Note Opinion and Credibility Use y_t

Opinion Equations

Factorize the opinion of expert $u_{1:t-1}$, into the product of the opinion about pose U_t times the opinion about texture V_t, B_t given pose.

$$p(u_t v_t b_t \,|\, u_{1:t-1} y_{1:t}) = p(v_t b_t \,|\, u_{1:t} y_{1:t}) p(u_t \,|\, u_{1:t-1} y_{1:t})$$

Opinion $=$ Texture Opinion $\times$ Pose Opinion

Texture opinions: The distribution of $V_t B_t$ given $u_{1:t} y_{1:t}$ is Gaussian with a mean and covariance that can be obtained using time dependent Kalman filter equations

$$Pcs(V_t B_t \,|\, u_{1:t} y_{1:t}) = Pcs(V_t B_t \,|\, u_{1:t-1} y_{1:t-1}) + c(u_t)' \Psi_w c(u_t)$$

$$E(V_t B_t \,|\, u_{1:t} y_{1:t}) = \frac{Pcs(V_t B_t \,|\, u_{1:t-1} y_{1:t-1}) E(V_t B_t \,|\, u_{1:t-1} y_{1:t-1}) + c(u_{t-1})' \Psi_w y_{t-1}}{Pcs(V_t B_t \,|\, u_{1:t-1} y_{1:t-1}) + c(u_t)' \Psi_w c(u_t)}$$

Note $E(V_t B_t \,|\, u_{1:t} y_{1:t})$ contains texture maps for object and background. $Var(V_t B_t \,|\, u_{1:t} y_{1:t})$ keeps the uncertainty about these maps.

Pose opinions: No analytical solution to distribution of u_t. However we can find most probable opinion u_t and approximate distribution using a Gaussian bump about that point. Note

$$p(u_t \mid u_{1:t-1} y_{1:t}) \propto p(u_t \mid u_{t-1}) p(y_t \mid u_{1:t} y_{1:t-1})$$

where

$$p(y_t \mid u_{1:t} y_{1:t-1}) = \int p(v_t b_t \mid u_{1:t-1} y_{1:t-1}) p(y_t \mid u_t v_t b_t) dv_t db_t$$

$$\hat{u}_t(u_{1:t-1}) = \operatorname*{argmax}_{u_t} p(u_t \mid u_{1:t-1} y_{1:t}) = \operatorname*{argmax}_{u_t} L(u_t, u_{1:t-1})$$

$$\begin{aligned} L(u_t, u_{1:t-1}) = \\ & -\frac{1}{2} \sum_{i \in \mathcal{O}(u_t)} \left(\frac{(y_t(i) - \mu_v(u_{1:t}, i))^2}{\sigma_v(u_{1:t}, i) + \sigma_w} - \frac{(y_t(i) - \mu_b(u_{1:t}, i))^2}{\sigma_b(u_{1:t}, i) + \sigma_w} \right) \\ & + \log p(u_t \mid u_{t-1}) \end{aligned}$$

$\hat{u}_t(u_{1:t-1})$ can be found very quickly using a Gauss-Newton method. The inverse Hessian $\hat{\sigma}_t(u_{1:t-1})$ also falls out easily from the Gauss-Newton method. The posterior distribution can then be approximated as a Gaussian $\phi(\cdot \mid \hat{u}_t(u_{1:t-1}), \hat{\sigma}_t(u_{1:t-1}))$ centered at $\hat{u}_t(u_{1:t-1})$ and with covariance $\hat{\sigma}_t(u_{1:t-1})$.

We can do importance sampling with $\phi(\cdot \mid \hat{u}_t(u_{1:t-1}), \alpha\hat{\sigma}_t(u_{1:t-1}))$ where $\alpha > 0$. As $\alpha \to 0$ same as picking the maximum.

Optic Flow as a Special Case: Suppose $p(u_t \mid u_{t+1})$ is uninformative, the background is a white noise process, i.e. $\sigma_b(u_t, i) \to \infty$ for all t, i and by time $t-2$ we are completely uncertain about the object texture, i.e.

$$Var(V_{t-1} \mid u_{1:t-2} y_{1:t-2}) \to \infty$$

It follows that

$$E(V_t \mid u_{1:t-1} y_{1:t-1}) = a_v(u_{t-1}) y_{t-1}$$

Thus

$$\operatorname*{argmax}_{u_t} \; p(u_t \,|\, u_{t-1} y_{1:t}) =$$

$$= \operatorname*{argmin}_{u_t} \sum_{i \in \mathcal{O}(u_t)} \frac{\left(y_t(i) - a_v(u_{t-1}) y_{t-1}(i)\right)^2}{\sigma_v(u_t, i) + \sigma_w}$$

The most probable u_t is that which minimizes the mismatch between the image pixels rendered by the object at time $t-1$ and the image at y_t shifted according to u_t. The Lucas-Kanade optic flow algorithm is simply the Newton-Gauss method as applied to minimize this error function.

Template matching as a Special Case: If $p(u_t \,|\, u_{t-1})$ is uninformative, the background is a white noise process and by time $t-2$ we are certain about the object texture map, i.e., $Var(V_{t-1} \,|\, u_{1:t-2} y_{1:t-2}) = 0$, then

$$E(V_t \,|\, u_{1:t-1} y_{1:t-1}) = E(V_t \,|\, u_{1:t-2} y_{1:t-2})$$

$$\operatorname*{argmax}_{u_t} \; p(u_t \,|\, u_{t-1} y_{1:t}) = \operatorname*{argmin}_{u_t} \sum_{i \in \mathcal{O}(u_t)} \frac{(y_t(i) - \mu_v(u_t, i))^2}{\sigma_w}$$

where $\mu_v(u_t, i)$ is the fixed object template, shifted by u_t.

Filter Distribution $=$ Sum Expert Opinion $\times$ Expert Credibility

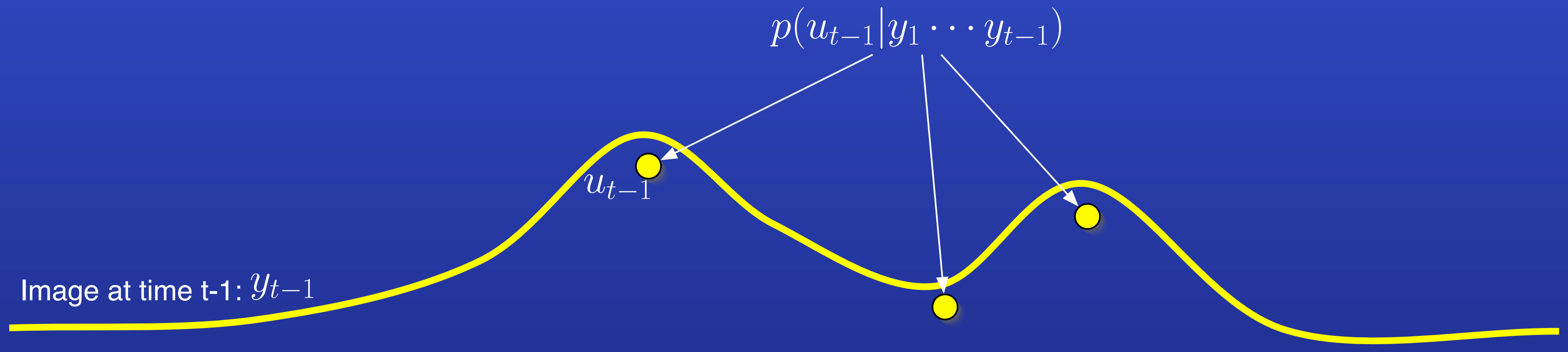
Credibility Equations

$$p(u_{1:t-1} \,|\, y_{1:t}) \propto p(u_{1:t-1} \,|\, y_{1:t-1}) p(y_t \,|\, u_{1:t-1} y_{1:t-1})$$

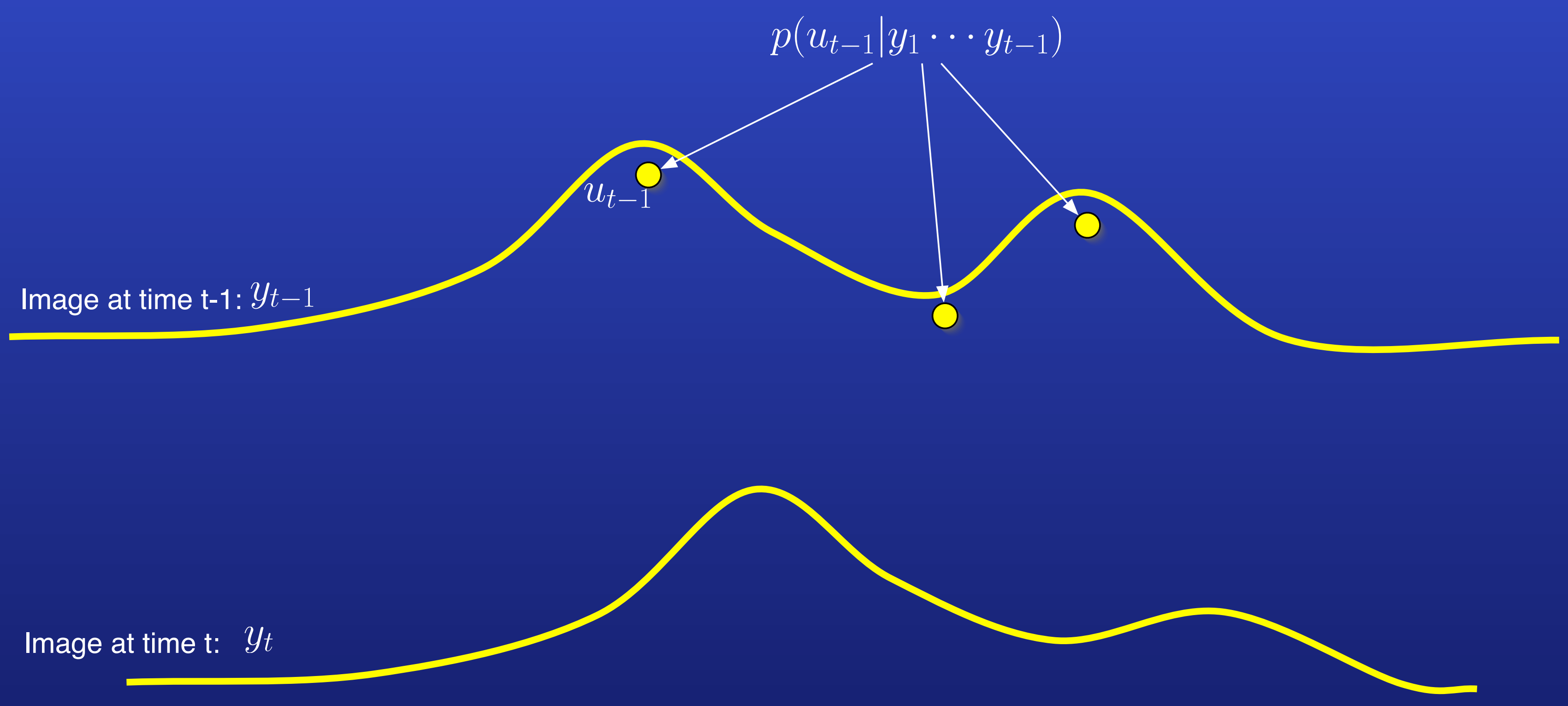
where

$$\begin{aligned} p(y_t \,|\, u_{1:t-1} y_{1:t-1}) &= \int p(y_t u_t \,|\, u_{1:t-1} y_{1:t-1}) du_t \\ &= \int p(y_t \,|\, u_{1:t} y_{1:t-1}) p(u_t \,|\, u_{t-1}) du_t \\ &\approx \sum_{i=1}^{s} w_t(u_{1:t-1}, i) \end{aligned}$$

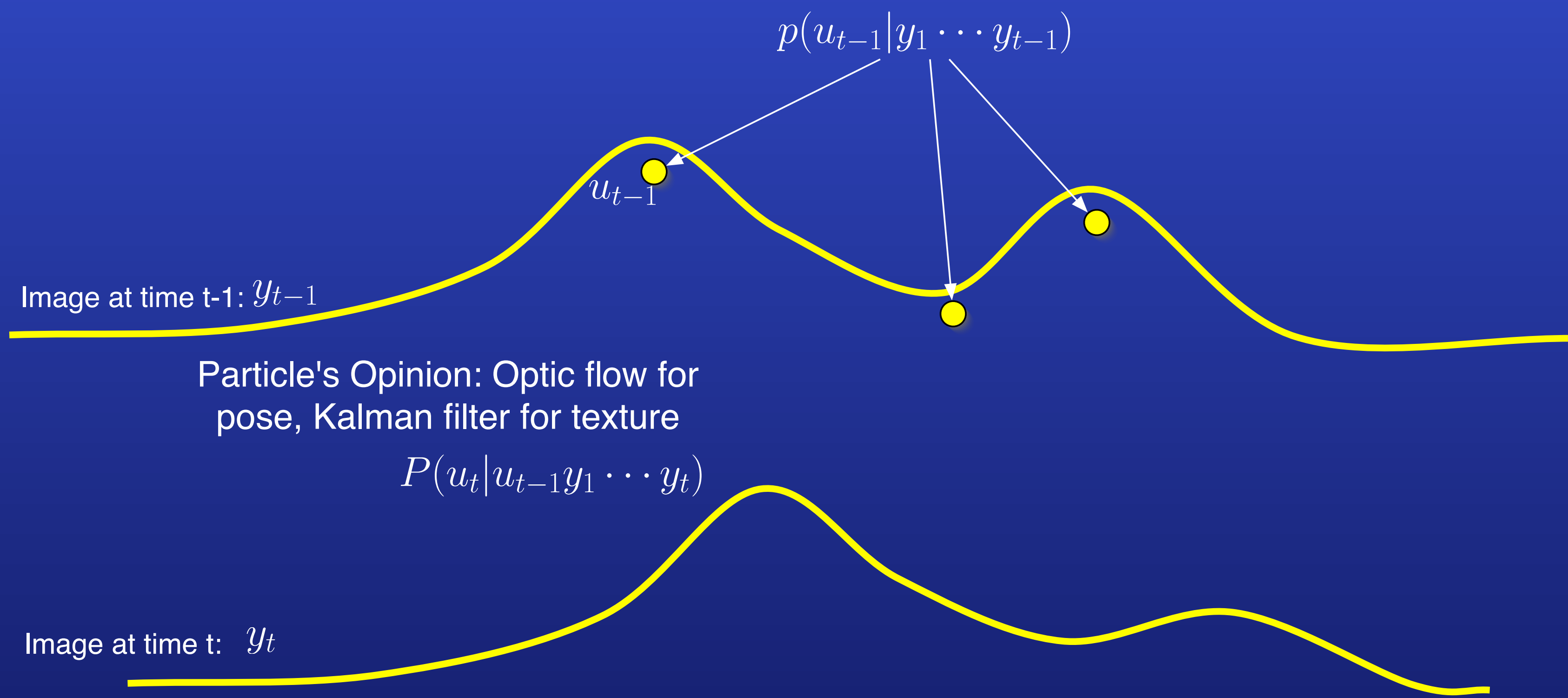
If $\alpha \longrightarrow 0$ simply get $\exp(L(\hat{u}_t, u_{1:t-1}))$.

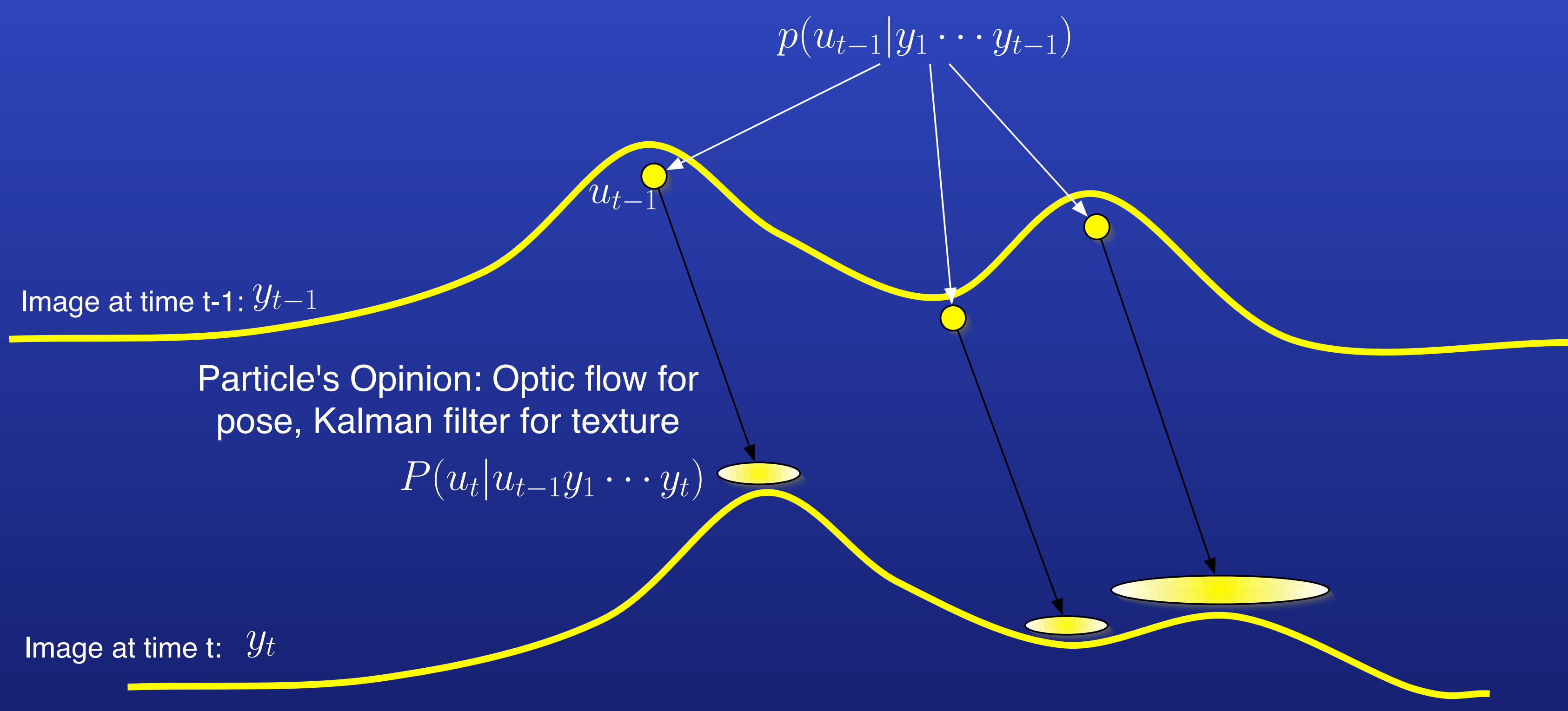


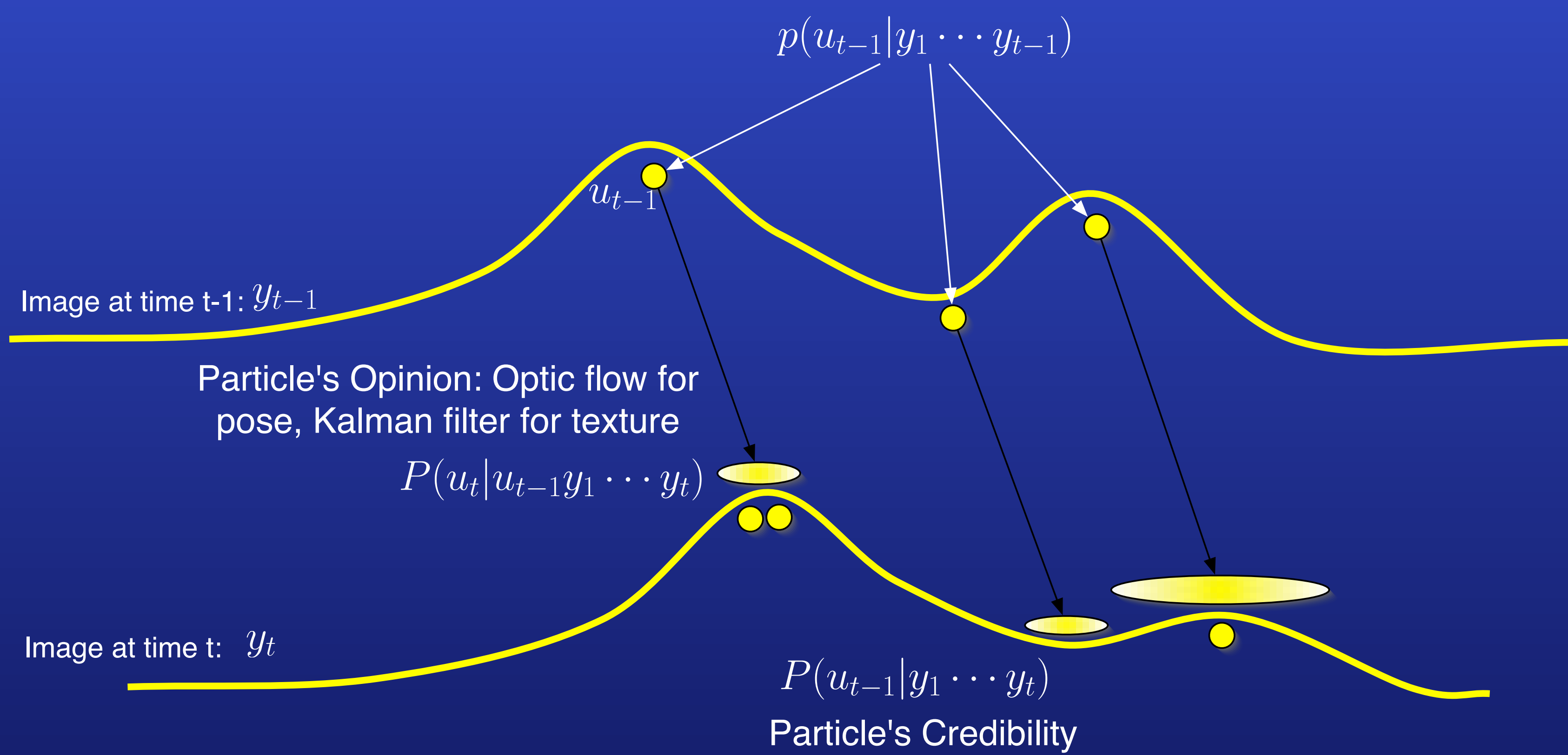
$p(u_{t-1}|y_1 \cdots y_{t-1})$
u_{t-1}
Image at time t-1: y_{t-1}

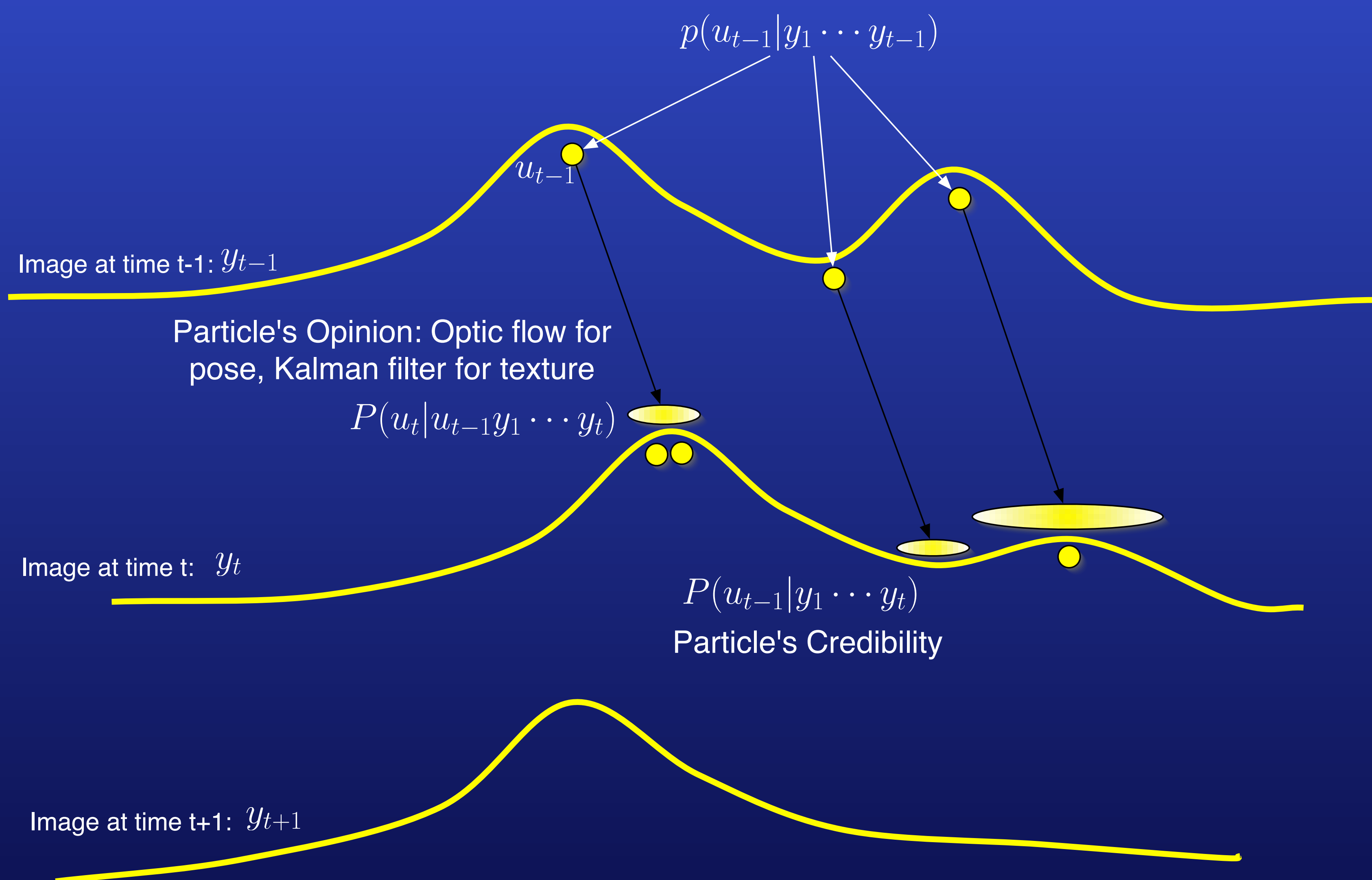


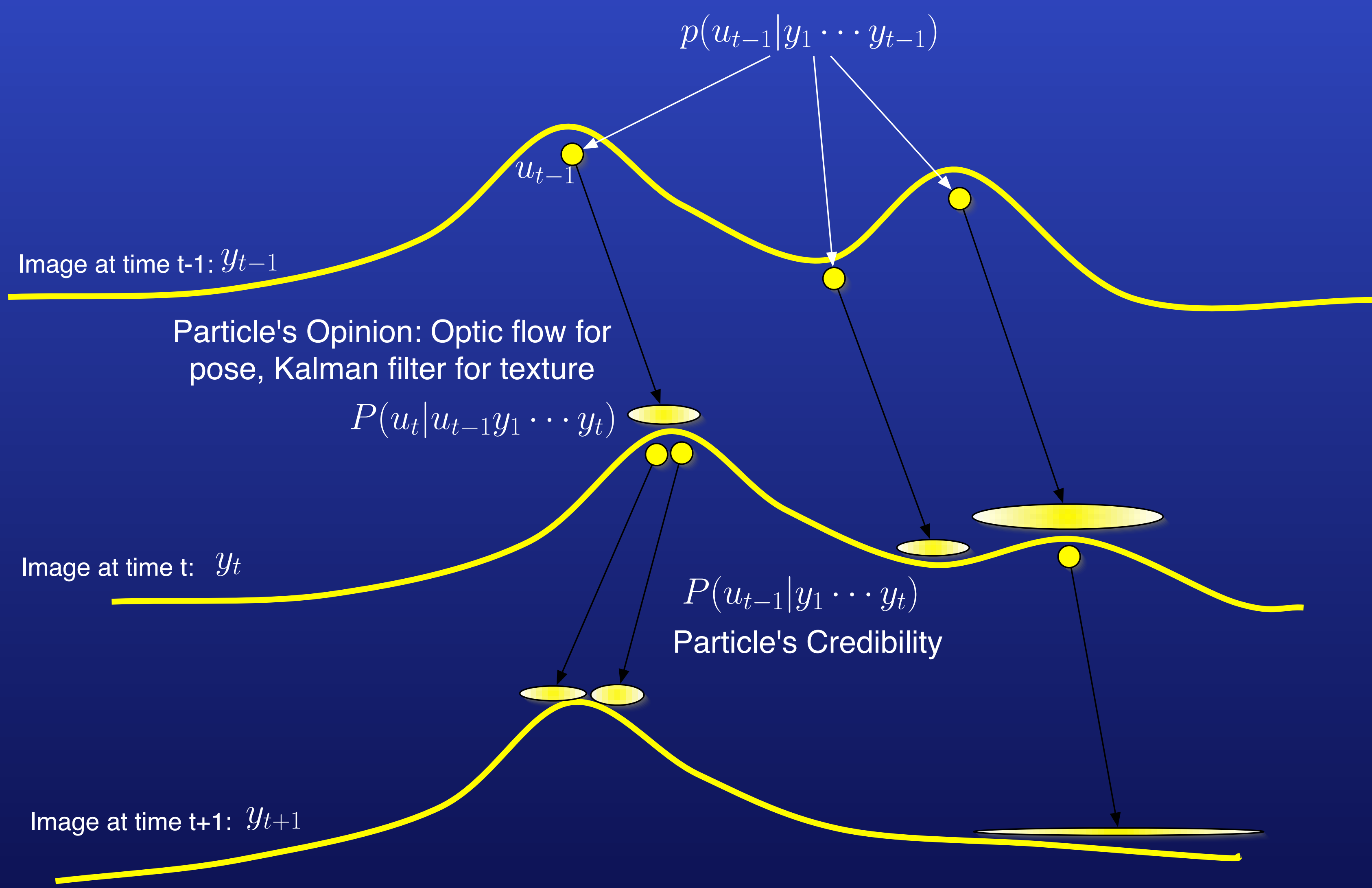
$p(u_{t-1}|y_1 \cdots y_{t-1})$
u_{t-1}
Image at time t-1: y_{t-1}
Image at time t: y_t

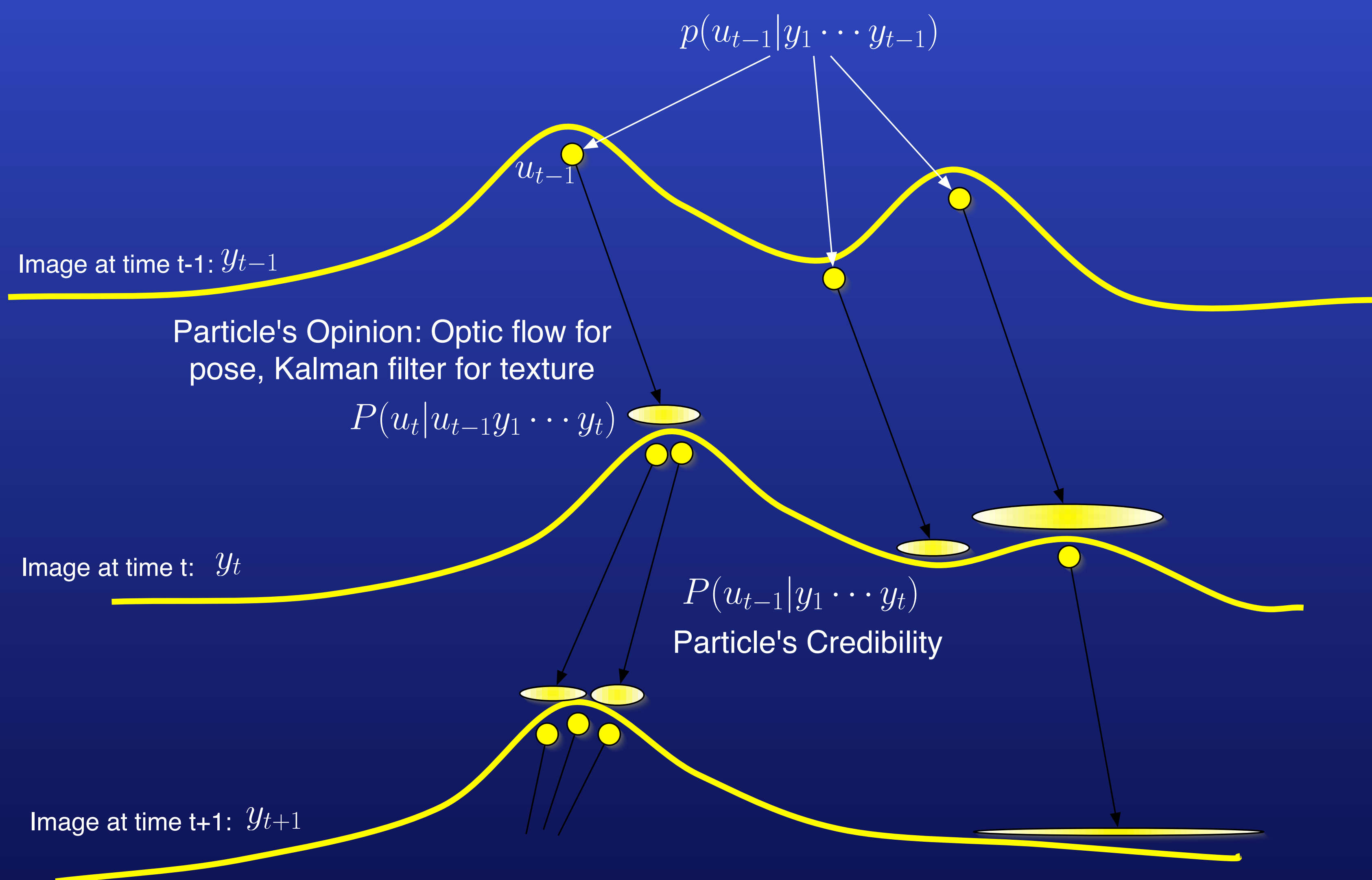

$p(u_{t-1}|y_1 \cdots y_{t-1})$
u_{t-1}
Image at time t-1: y_{t-1}
Particle's Opinion: Optic flow for pose, Kalman filter for texture
$P(u_t|u_{t-1}y_1 \cdots y_t)$
Image at time t: y_t


$p(u_{t-1}|y_1 \cdots y_{t-1})$
u_{t-1}
Image at time t-1: y_{t-1}
Particle's Opinion: Optic flow for pose, Kalman filter for texture
$P(u_t|u_{t-1}y_1 \cdots y_t)$
Image at time t: y_t


$p(u_{t-1}|y_1 \cdots y_{t-1})$
u_{t-1}
Image at time t-1: y_{t-1}
Particle's Opinion: Optic flow for pose, Kalman filter for texture
$P(u_t|u_{t-1}y_1 \cdots y_t)$
Image at time t: y_t
$P(u_{t-1}|y_1 \cdots y_t)$
Particle's Credibility


$p(u_{t-1}|y_1 \cdots y_{t-1})$
u_{t-1}
Image at time t-1: y_{t-1}
Particle's Opinion: Optic flow for pose, Kalman filter for texture
$P(u_t|u_{t-1}y_1 \cdots y_t)$
Image at time t: y_t
$P(u_{t-1}|y_1 \cdots y_t)$
Particle's Credibility
Image at time t+1: y_{t+1}


$p(u_{t-1}|y_1 \cdots y_{t-1})$
u_{t-1}
Image at time t-1: y_{t-1}
Particle's Opinion: Optic flow for pose, Kalman filter for texture
$P(u_t|u_{t-1}y_1 \cdots y_t)$
Image at time t: y_t
$P(u_{t-1}|y_1 \cdots y_t)$
Particle's Credibility
Image at time t+1: y_{t+1}



$p(u_{t-1}|y_1 \cdots y_{t-1})$
u_{t-1}
Image at time t-1: y_{t-1}
Particle's Opinion: Optic flow for pose, Kalman filter for texture
$P(u_t|u_{t-1}y_1 \cdots y_t)$
Image at time t: y_t
$P(u_{t-1}|y_1 \cdots y_t)$
Particle's Credibility
Image at time t+1: y_{t+1}

Example

Conclusions

- 3D Tracking can be casted as a Conditionally Gaussian Filtering Problem.
- Optic-Flow-Like algorithm provides most probable pose at time t given the images up to time t and the poses up to time $t-1$.
- This avoids needle-in-haystack problem.
- Object and Background texture distribution is learned via Kalman filters.

- Optic-Flow and template matches emerge as special cases of optimal inference under some conditions.
- In practice optimal inference behaves as a combination of motion-like and template-like tracking.

MENU