Object Tracking with Bayesian Estimation of Dynamic Layer Representation

(H. Tao, H. Sawhney and R. Kumar)

PAMI, Jan 2002

CVPR 2000

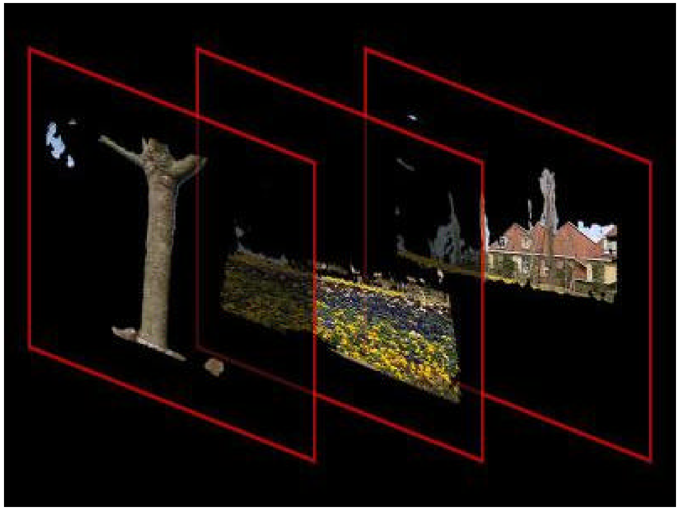
Presented by: Adeel Bhutta

Logistique

- Introduction
  - Basic idea
  - MAP Estimation
  - Notation
- Layer Representation
  - Technique
  - Applications and Results
- Implementation
- Discussion

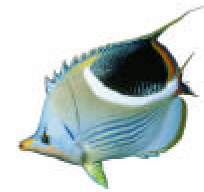
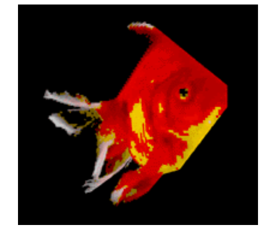
Basic Idea

- Object Tracking by Layer Representation
  - **Object Tracking**
    - Estimation of complete representation of foreground and background objects over time
  - **Layer Representation**
    - Region of homogeneous motion in an image sequence

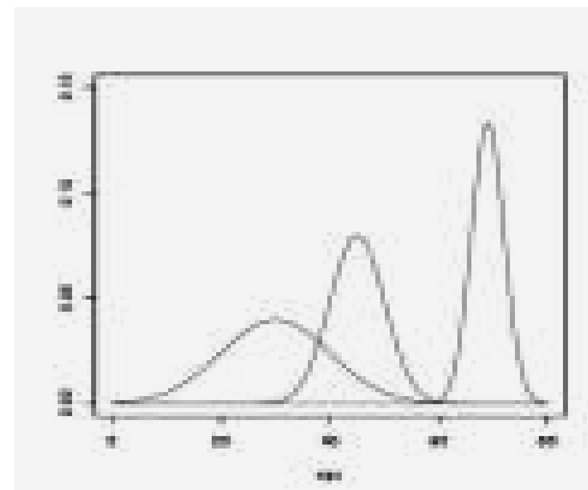


MAP Estimation

- 3 classes of fish
- 100 pre-classified fish (50-10-40)
- Feature is … Mass
- Model of Class Feature
  - Gaussian … mean, variance for each class
- New Observation (mass of fish)
- Find which class it belongs to using MAP Estimation
- Mass of Fish (D), Class (h_i)
- Find P (h_i | D) ?

20g 10g 25g



$$P(h_i \mid D) = \frac{P(D \mid h_i)P(h_i)}{P(D)}$$

$$f(D \mid m, s^2) = \frac{1}{\sqrt{2ps^2}} e^{\frac{-(D-m)^2}{(2s^2)}}$$

MAP Estimation

Likelihood

Prior

$$P(h_i \mid D) = \frac{P(D \mid h_i)P(h_i)}{P(D)}$$

Posterior

- $P(D \mid h_i)$: Probability of observing Data 'D' given 'h' (**Likelihood**)
- $P(h_i)$: A priori Probability of particular hypothesis 'h' before knowing "D"
- $P(h_i \mid D)$: Posterior hypothesis given observation (**Posterior**)
- $P(D)$: Probability of observation

- Find 'i' that maximizes the posterior probability (MAP)

$$h_{MAP} = \arg\max_{h \in H} P(D|h) \cdot P(h)$$

- Recap:
  - Have classes with priors
  - Choose features, model them
  - Find Max posterior probability for new observation

Problem Formulation - I

- Multi-object tracking as a 2D motion layer estimation problem with a view towards achieving completeness of representation

- Goal:
  - Layers (support) + Motion models
  - Maintain coherency between motion, appearance, and shape of each layer over time (***Main Contribution***)

- Estimate Layers with Maximum posteriori probability using Generalized EM algorithm

Problem Model

- New Observation: I_t (*remember new fish*)
- Classes: Layers
- Features: (*remember mass*)
  - Shape of Layer: Φ_t
  - Motion of Layer: Θ_t
  - Appearance of Layer: A_t
- State of Layer: $\Lambda_t = (\Phi_t, \Theta_t, A_t)$
  - Complete representation (Big Claim!)

Problem Formulation - II

- Dynamic Layer Estimation

$$\max_{\Lambda_t} \arg P(\Lambda_t \mid I_t, \ldots, I_0, \Lambda_{t-1}, \ldots, \Lambda_0)$$

- Markovian Assumption:
  - Parameters at current time instant depend only on those at the previous time instant.

$$\max_{\Lambda_t} \arg P(\Lambda_t \mid I_t, \ldots, I_0, \Lambda_{t-1}, \ldots, \Lambda_0) = \max_{\Lambda_t} \arg P(\Lambda_t \mid I_t, I_{t-1}, \Lambda_{t-1})$$

- Bayes' Rule

$$\max_{\Lambda_t} \arg P(\Lambda_t \mid I_t, \ldots, I_0, \Lambda_{t-1}, \ldots, \Lambda_0) \propto \max_{\Lambda_t} \arg P(I_t \mid \Lambda_t, I_{t-1}, \Lambda_{t-1}) P(\Lambda_t \mid I_{t-1}, \Lambda_{t-1})$$

Motion Model ($\Theta_{t,j}$)

- Foreground $\dot{m}_{t,j}$ $\dot{w}_{t,j}$
  - 2D Rigid Motion: Translation + Rotation
  - Constant Velocity Model
    - *Vehicles move with relatively constant speed*
- Background
  - Planar Projective
    - *Good for aerial videos*
- Motion (*modeled as Gaussian distribution*)

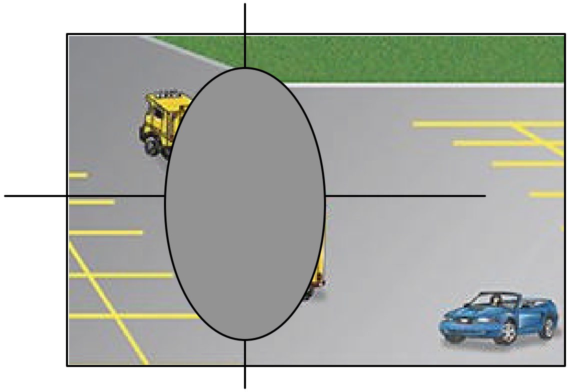
$$P(\Theta_{t,j} \mid \Theta_{t-1,j}) = N(\Theta_{t,j} : \Theta_{t-1,j}, diag[s_m^2, s_m^2, s_w^2])$$

Dynamic Segmentation Prior

- Shape Prior Parameters $\Phi_t = \{l_t, s_t\}$

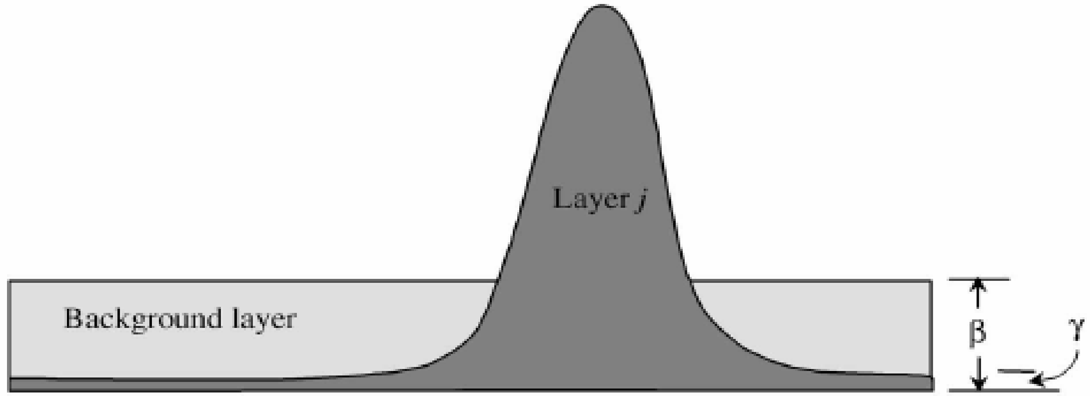
- Dynamics of shape prior (changes over time)
  - Constancy of shape over time
  - Modeled with Gaussian

$$P(\Phi_{t,j} \mid \Phi_{t-1,j}) = N(\Phi_{t,j} : \Phi_{t-1,j}, diag[s_{ls}^{2}, s_{ls}^{2}])$$



Dynamic Segmentation Prior

- Goal: Assign pixels to layers
  - Background: Uniform Prior
  - Foreground: Gaussian segmentation prior (elliptical)



- Probability of a pixel location (x_i) belonging to certain layer 'j'

$$L_{t,j}(x_i) = \begin{cases} g + \exp[-(x_i - m_{t,j})^T \Sigma_{t,j}^{-1} (x_i - m_{t,j})/2] \\ b \end{cases}$$

- $j >= 1$
- $j = 0$

$\mathbf{x_i}$: Image coordinates of ith pixel

g : uncertainty of layer shape (non-elliptical)

Dynamic Segmentation Prior

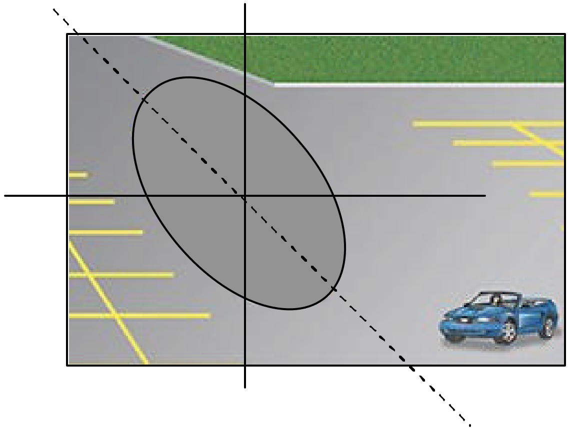
- Covariance Matrix:

$$\sum_{t,j} = R^T(-w_{t,j})\, diag[l_{t,j}^2, s_{t,j}^2]\, R(-w_{t,j})$$

$l_{t,j}$, $s_{t,j}$: proportional to length of major or minor axis of contours

- Normalize the priors

$$S_t(x_i) = L_{t,j}(x_i) / \sum_{j=0}^{g-1} L_{t,j}(x_i)$$



Coordinate Transformation

- Coordinate Transformation from Original to local coordinate system (compensating the motion)

$$x_i^j = R(-w_j)(x_i - m_i)$$

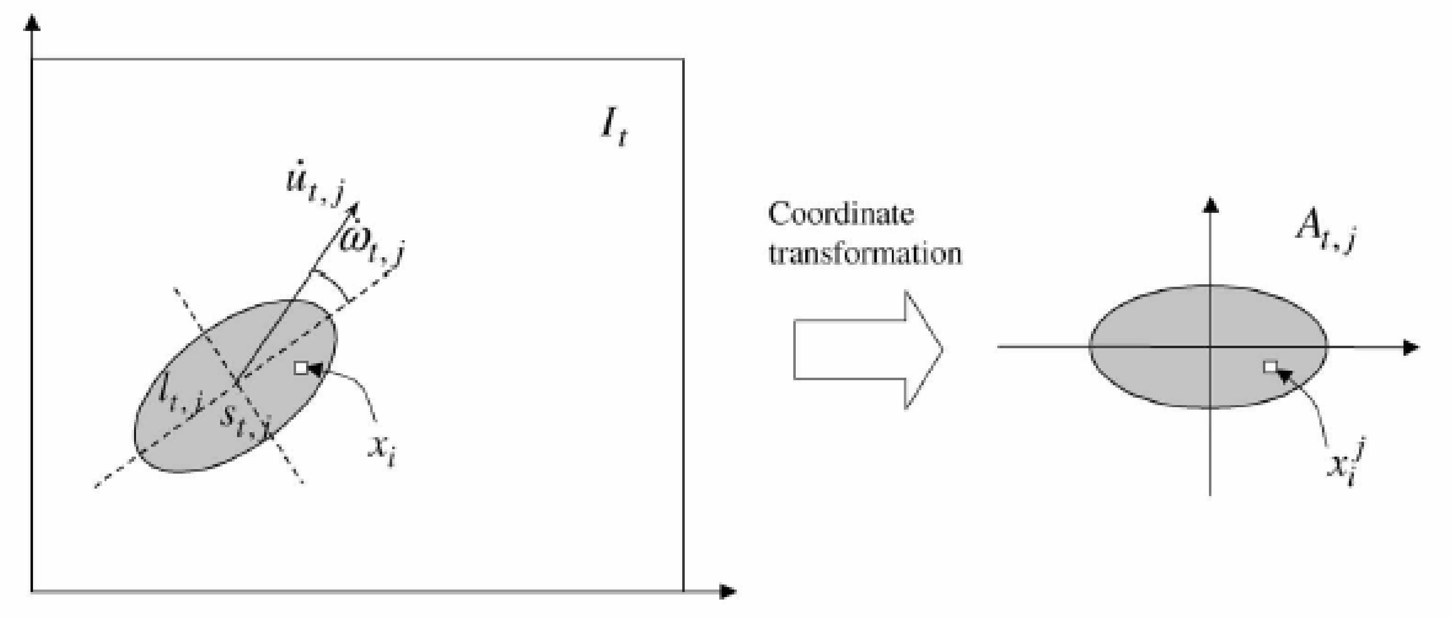


Image Observation Model & Layer Appearance Model

- Observation for layer ‘j’

$$P(I_t(x_i) \mid A_{t,j}(x_i^j)) = N(I_t(x_i) : A_{t,j}(x_i^j), s_I^2)$$

  - What is the probability that estimate of pixel intensity in some layer is seen in new Image

- Intensity value of a pixel in layer ‘j’

$$P(A_{t,j}(x_i^j) \mid A_{t-1,j}(x_i^j)) = N(A_{t,j}(x_i^j) : A_{t-1,j}(x_i^j), s_A^2)$$

  - What is the change in pixel intensity if we know the intensity in last image.

Update

Summary of Priors

Likelihood

- Motion

$$P(\Theta_{t,j} \mid \Theta_{t-1,j}) = N(\Theta_{t,j} : \Theta_{t-1,j}, diag[s_m^2, s_m^2, s_w^2])$$

- Shape (Priors and Dynamics)

$$S_t(x_i) = L_{t,j}(x_i) / \sum_{j=0}^{g-1} L_{t,j}(x_i)$$

$$P(\Phi_{t,j} \mid \Phi_{t-1,j}) = N(\Phi_{t,j} : \Phi_{t-1,j}, diag[s_{ls}^2, s_{ls}^2])$$

- Appearance (Observation + Intensity Value)

$$P(I_t(x_i) \mid A_{t,j}(x_i^j)) = N(I_t(x_i) : A_{t,j}(x_i^j), s_I^2)$$

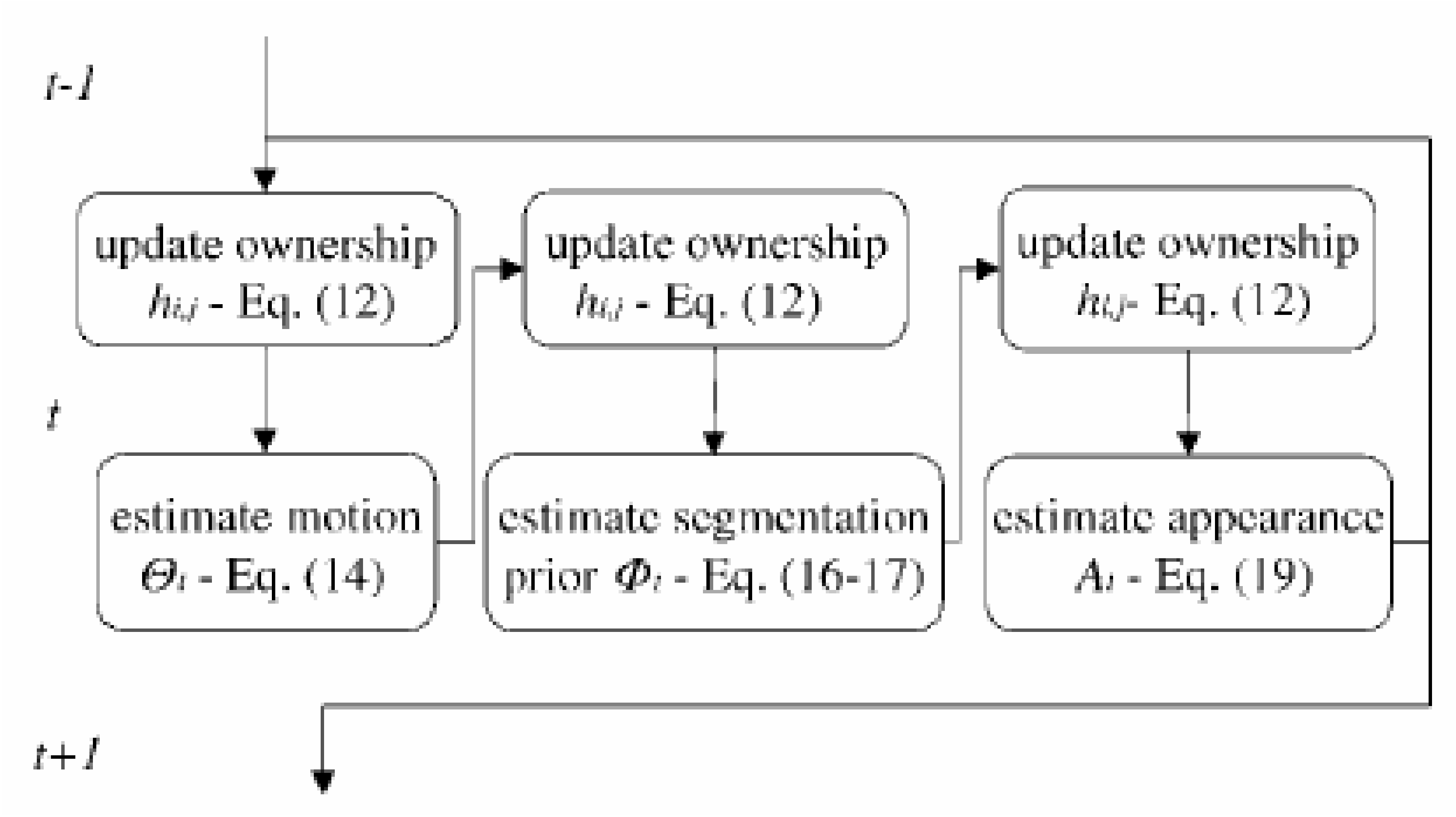
$$P(A_{t,j}(x_i^j) \mid A_{t-1,j}(x_i^j)) = N(A_{t,j}(x_i^j) : A_{t-1,j}(x_i^j), s_A^2)$$

EM Algorithm

$$\max_{\Lambda t} \arg P(I_t \mid \Lambda_t, I_{t-1}, \Lambda_{t-1}) P(\Lambda_t \mid I_{t-1}, \Lambda_{t-1})$$

- For every time instant
  - Calculate Segmentation (Expectation step)
  - Update layer parameters (Maximization step)
- Questions:
  - Correspondence between pixels and layers
  - Computation of optimal layer parameters
- Use generalized EM algorithm to iteratively optimize.

EM Algorithm



Optimization

- Layer Ownership

Shape Prior

$$
\begin{aligned}
h_{i,j} &= P(z_t(x_i) = j | I_t, \Lambda'_t, \Lambda_{t-1}, I_{t-1}) \\
&= \frac{P(I_t | z_t(x_i) = j, \Lambda'_t, \Lambda_{t-1}, I_{t-1}) P(z_t(x_i) = j | \Lambda'_t, \Lambda_{t-1}, I_{t-1})}{P(I_t | \Lambda'_t, \Lambda_{t-1}, I_{t-1})} \\
&= P(I_t(x_i) | A'_{t,j}(x_i^j)) S_{t,j}(x_i) / Z.
\end{aligned}
$$

Constant

Appearance

- z_t: hidden variable indicating association of each pixel to each layer
- **A':** Appearance from previous iteration.

Eq. 12

Optimization

- Motion Estimation

$$\sum_{j=1}^{g-1} \log N\left(\Theta_{t,j} : \Theta_{t-1,j}, diag\left[\sigma_\mu^2, \sigma_\mu^2, \sigma_\omega^2\right]\right) +$$

$$\sum_{i=0}^{n-1}\sum_{j=1}^{g-1} h_{i,j}\left\{\log S_{t,j}(x_i) + \log P\left(I_t(x_i)|A_{t,j}(x_i^j)\right)\right\}.$$

$$\min_{\Theta_{t,j}} \arg |\mu_{t,j} - \mu_{t-1,j}|/\sigma_\mu^2 + |\omega_{t,j} - \omega_{t-1,j}|/\sigma_\omega^2 -$$

Motion

$$\sum_{i=0}^{n-1} 2h_{i,j} \log S_{t,j}(x_i) + \sum_{i=0}^{n-1} h_{i,j}\left(I_t(x_i) - A_{t,j}(x_i^j)\right)^2/\sigma_I^2.$$

Shape

SSD

- **Solution obtained by searching in space of translation and rotation**

Eq. 14

Optimization

- Shape Estimation

$$\max_{\Phi_t} \arg \; f = \sum_{j=0}^{g-1} \log N\left(\Phi_{t,j} : \Phi_{t-1,j}, diag\left[\sigma_{ls}^2, \sigma_{ls}^2\right]\right) + \sum_{i=0}^{n-1} \sum_{j=0}^{g-1} h_{i,j} \log S_{t,j}(x_i).$$

$$\frac{\partial f}{\partial s_{t,j}} = \sum_{i=0}^{n-1} \frac{h_{i,j}\left(D(x_i) - L_{t,j}(x_i)\right)}{L_{t,j}(x_i) D(x_i)} (L_{t,j}(x_i) - \gamma) y_{i,j,y}^2 / s_{t,j}^3 - (s_{t,j} - s_{t-1,j}) / \sigma_{ls}^2,$$

$$\frac{\partial f}{\partial l_{t,j}} = \sum_{i=0}^{n-1} \frac{h_{i,j}(D(x_i) - L_{t,j}(x_i))}{L_{t,j}(x_i) D(x_i)} (L_{t,j}(x_i) - \gamma) y_{i,j,x}^2 / l_{t,j}^3 - (l_{t,j} - l_{t-1,j}) / \sigma_{ls}^2$$

Ref: Appendix C **Eq. 16-17**

Optimization

- Appearance Estimation

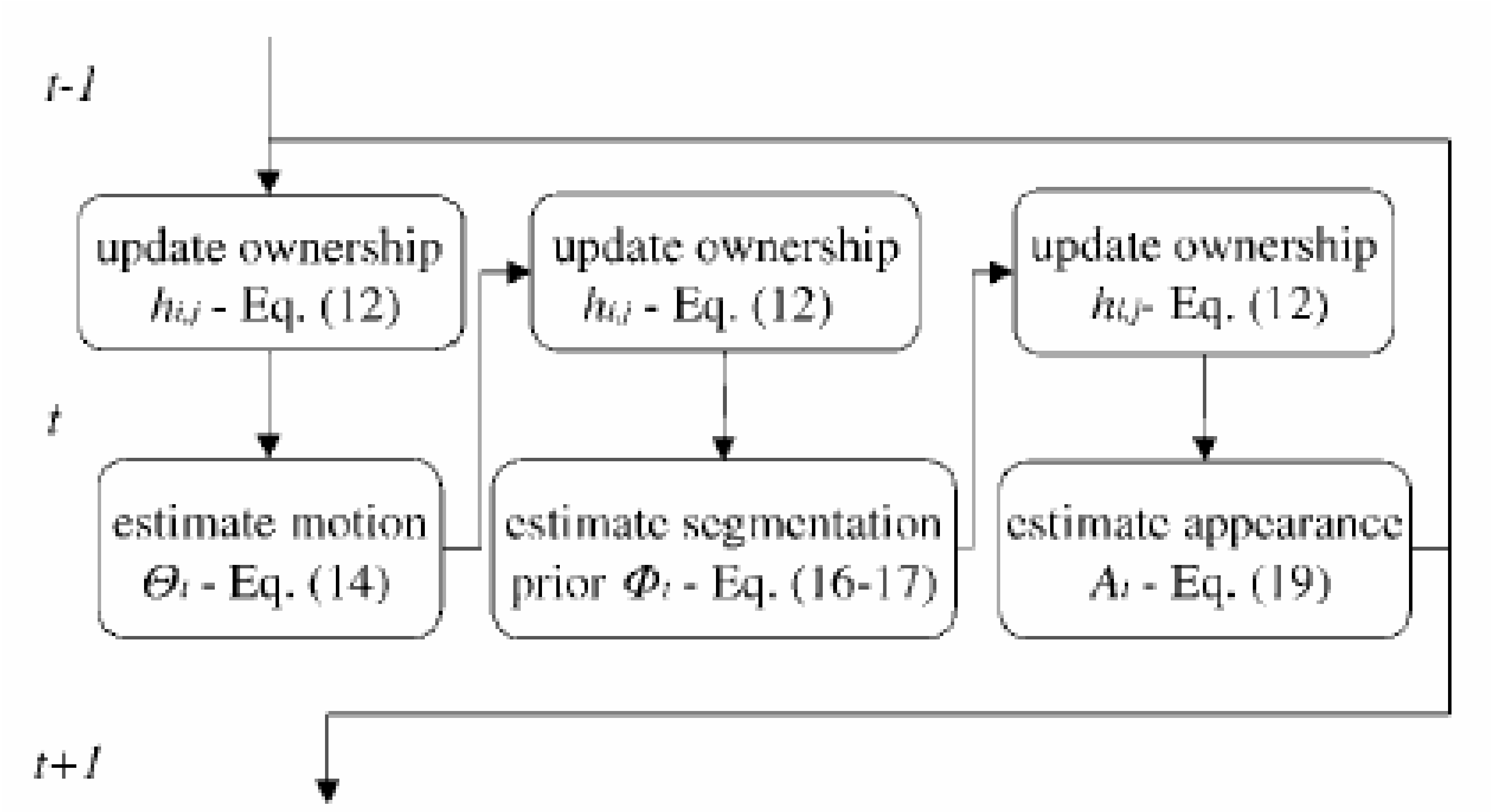
$$\max_{A_{t,j}} \arg \sum_{i=0}^{n-1} \Big\{ \log\Big(N\big(A_{t,j}(x_i^j) : A_{t-1,j}(x_i^j), \sigma_A^2\big)\Big) + h_{i,j} \log P\Big(I_t(x_i) | A_{t,j}(x_i^j)\Big)\Big\}.$$

$$A_{t,j}(x_i^j) = \frac{A_{t-1,j}(x_i^j)/\sigma_A^2 + h_{i,j} I_t(x_i)/\sigma_I^2}{(1/\sigma_A^2 + h_{i,j}/\sigma_I^2)}.$$

Note: Error in Eq. 19

Ref: Appendix D **Eq. 19**

EM Algorithm



Results – Vehicle Turning

Results – Vehicle Stop

Results – Vehicle Passing

Results – Vehicle Passing

Implementation

1. Acquire the Video
2. Find Registration parameters (Ref: [10])
3. Initialization
    1. Find Change Blob
    2. Implement State Machine (Fig. 7)
4. Tracking (EM Algorithm, Fig. 4)
    1. Multiple iterations if needed
5. Repeat step 2-5 for each new image

Discussion - Strengths

- Motion, Shape, and Appearance Models used.
- Competition between Layers for Ownership of each pixel (Robustness)
- Examples:
  - Can track object which ‘stopped’ (Appearance)
  - Can track close objects with similar motion (Appearance and shape priors)
  - Can track objects that change shape (!) (Appearance and Layer Ownership)

Discussion - Weakness

- Will work for rigid objects only!
- Video should be from camera far off
- Everything is gaussian!
- Overly complicated approach to solve tracking problem!

Discussion - Problems

- Stop Sequence
  - Missed 2nd Vehicle Entry altogether
  - Tracking even when box missing
- Turning Sequence
  - Slightly track corner vehicles (camera motion)
- Passing Vehicle
  - Takes long time to start tracking

Discussion - Ideas

- Q: How to incorporate complicated segmentation priors (Non-Rigid Objects)?
- Q: Occlusion?

Useful Links

- [10]: Berger et. al.,
- [17]: Tao’s website: http://www.cse.ucsc.edu/~tao/LAYER/