A Background Layer Model for Object Tracking through Occlusion

Yue Zhou and Hai Tao

UC-Santa Cruz

Presented by: Mei-Fang Huang

CSE252C class, November 6, 2003

Overview

- Object tracking problems
- Dynamic layer model
- Estimating model parameters
- Implementation and results
- Conclusion and discussion

Tracking Problems

Goal:

Estimate 2D or 3D positions of foreground and background objects over time

Approach:

- Model-based
  - *e.g.* ***Condensation***
- Layer-based
  - *e.g.* ***Dynamic layer model***
  - ***e.g. Flexible sprites***

Dynamic Layer Model

- Layer model
  - Represent of moving objects with different motion into different layers
  - "Layered representation for motion analysis"

    -Wang & Adelson, 1993 CVPR
- Dynamic Layer Model
  - Dynamic ?
  - Components ?
  - Estimation ?

Dynamically update layer model
Allow layer order alteration and layer deletion and creation

Contributions of DLM

- "Object tracking with bayesian estimation of dynamic layer representation"
    - H. Tao, H. S. Sawhney, R. Kumar, 2002 IEEE transactions on pattern analysis and machine intelligence
  - Complete representation
  - Dynamic estimation
  - Insufficient in objects with ***occlusion***
- New ideas in this paper

• Introduce "Ordering Information" **– Z-depth**

-Foreground / background layer ordering

-Allow multiple background layers

New Layer Representation

- Scene representation with ***Depth Ordering Information***
- Moving objects are described as ***Foreground Layers***
- Others belong to ***Background Occluding Layers***
- Interlace foreground and background layers

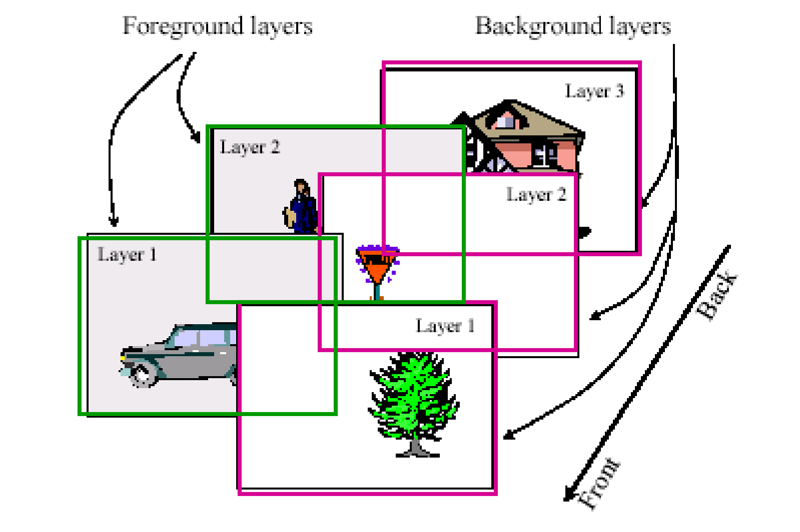
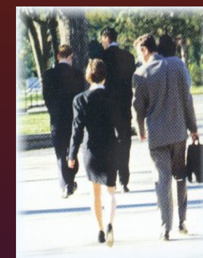


Figure 1. The ordered layer model.

Difference from previous work

- Use multiple background layers instead of only one background layer
- Try to solve complicated occlusion problems

How does it work?

Video frame(t-1)  Video frame(t) 

\+ \+

Layer Representation → **Estimation** → **Optimize object state**

Object state(t-1)

- Layer ordering
- Motion
- Shape
- Appearance

Object state(t)

- ?

Optimal object state(t)

- Layer ordering
- Motion
- Shape
- Appearance

Goal

We want to estimate:

- Foreground layer ordering
- Background layer
- Motion layer parameters

Achieved by:

- MAP framework (Maximum A Posteriori)

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

The MAP estimation

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

Λ_t : the state of the tracker at time t

I_t : the image observation at time t

The MAP estimation

Λ_t : the state of the tracker at time t

I_t : the image observation at time t

Using Bayes rule & HMM

$$P(\Lambda_t \mid \Lambda_{t-1}, I_t, \ldots, I_0) = P(I_t \mid \Lambda_t) \cdot P(\Lambda_t \mid \Lambda_{t-1})$$

$$P(Y \mid X, E) = \frac{P(X \mid Y, E) P(Y \mid E)}{P(X \mid E)}$$

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

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

$$= \arg\max_{\Lambda_t} P(I_t \mid \Lambda_t) \cdot P(\Lambda_t \mid \Lambda_{t-1})$$

Estimation

- Recall that we have

$$P(\Lambda_t \mid \Lambda_{t-1}, I_t, \ldots, I_0) = \underbrace{\boxed{P(I_t \mid \Lambda_t)}}_{\text{Likelihood}} \cdot \underbrace{\boxed{P(\Lambda_t \mid \Lambda_{t-1})}}_{\text{Prior}}$$

Apply the motion layer models

- Prior function

$$P(\Lambda_t \mid \Lambda_{t-1}) = P_{order} \cdot P_{fg\_shape} \cdot P_{bg\_shape} \cdot P_{motion} \cdot P_{appearance}$$

where

$$P_{order} = P(o_t \mid o_{t-1})$$

$$P_{fg\_shape} = \prod_{j=1}^{L} \prod_{i=1}^{N_j} P(\tau_{t,j}(x_i) \mid \tau_{t-1,j}(x_i))$$

$$P_{bg\_shape} = \prod_{j=1}^{L+1} \prod_{i=1}^{N_j} P(\pi_{t,j}(x_i) \mid \pi_{t-1,j}(x_i))$$

$$P_{motion} = \prod_{j=1}^{L} P(\theta_{t,j} \mid \theta_{t-1,j})$$

$$P_{appearance} = \prod_{j=1}^{L} \prod_{i=1}^{N_j} P(A_{t,j}(x_i) \mid A_{t-1,j}(x_i))$$

Apply the motion layer models

- Likelihood function

$$P(I_t \mid \Lambda_t) = \prod_{i=1}^{N} (P_{bgo}(x_i) + P_{fgo}(x_i))$$

- One background layer/ multiple foreground layer

Background observation probability

Probability in one background layer

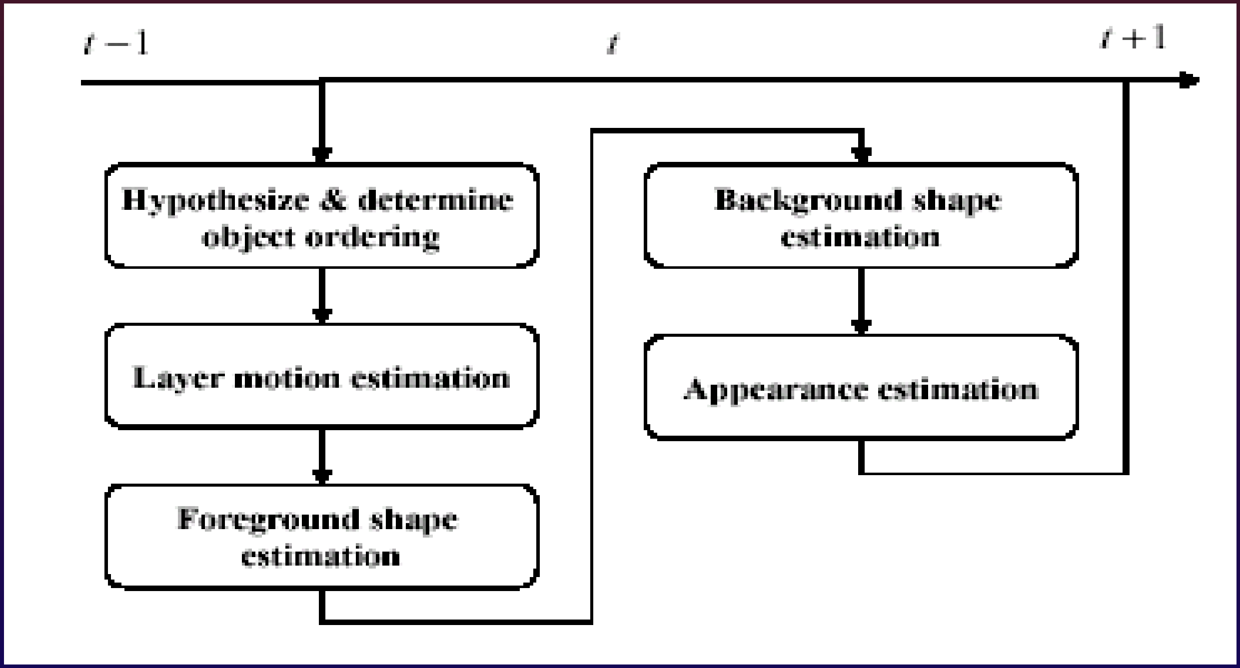
$$P_{bgo}(x_i) = P(I(x_i) \mid B(x_i)) \cdot P_B(x_i)$$

$$P_{fgo}(x_i) = \sum_{j=1}^{L} [P(I_j(x_i) \mid A_j(x_i)) \cdot P_j(x_i)]$$

jth foreground is visible

Estimate Sub-Problems

Approximate solution: Divide it into sub-problems



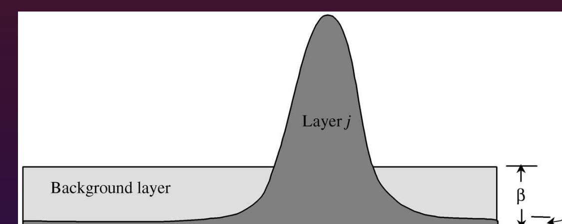
Motion Layer Analysis

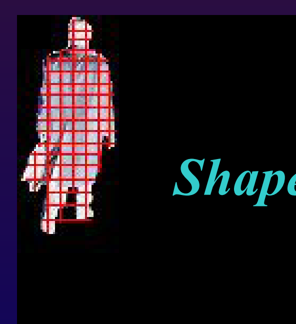
- Model the shape and appearance of each object
- Some approaches
  - Gaussian distributions
  - Markov Random Fields
  - Mixture models

    (Use EM to get weight)

Gaussian segmentation

prior function

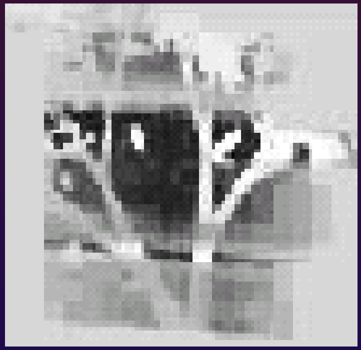




Shape map/mask

Motion Layer Parameters

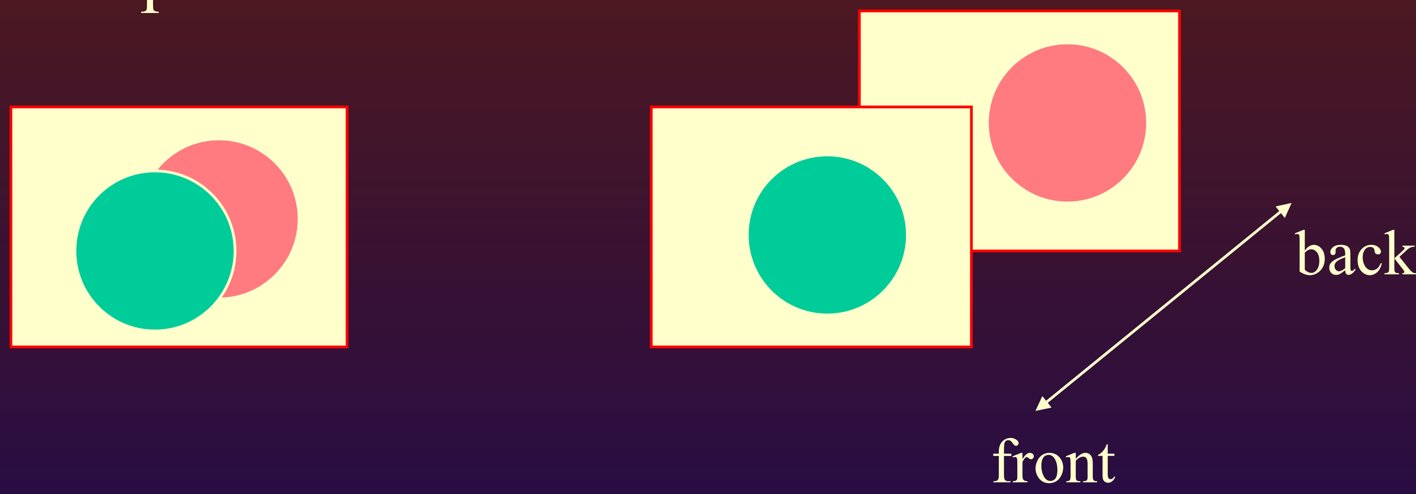
- Foreground Layer
  - Motion / Shape / Appearance models
  - Parameters: Position,
    Orientation,
    Scale
- Background Layer
  - Shape / Appearance models
  - Shared a ***single*** motion
- Depth ordering



Background Layer

What to do if there is occlusion?

- Depth order



Layer ordering → Layer model ⇒ Optimal model

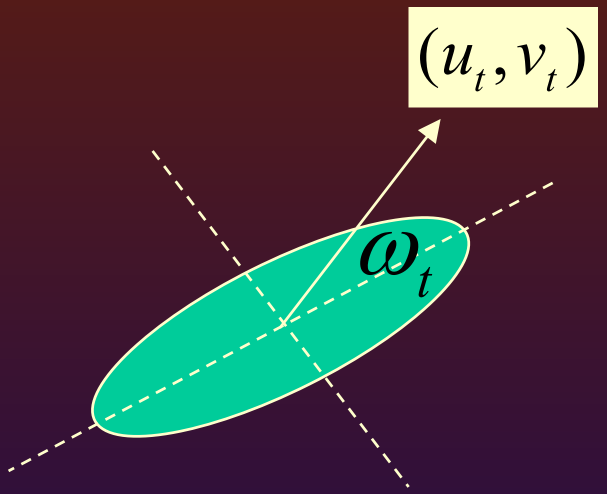
EM

Implementation of DLM

- Motion models
- Shape models
- Layer visibility
- Shape dynamics
- Appearance model

Motion Models

- Foreground
  - Position
  - Translation + rotation
  - Constant velocity
- Background
  - Planar projective



$$P(\theta_t \mid \theta_{t-1}) = N(\theta_t : \Phi\,\theta_{t-1}, Q)$$

Φ : standard transition matrix for a constant velocity model

$\Theta = [\mu, \omega, s, \dot{\mu}, \dot{\omega}, \dot{s}]$

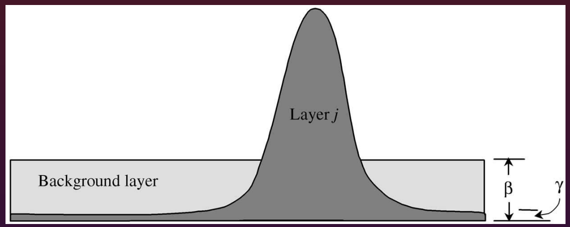
Implementation of DLM

- Motion models
- Shape models
- Layer visibility
- Shape dynamics
- Appearance model

Shape Models

- Shape map (a priori)
  - Foreground layers $\tau_{i,j}(x_i)$
  - Background layers $\pi_{i,j}(x_i)$

Gaussian segmentation prior function



- Assumption
  - Each pixel only belongs to ***ONE*** background layer

$$\sum_j \pi_{i,j}(x_i) = 1$$

Implementation of DLM

- Motion models
- Shape models
- Layer visibility
- Shape dynamics
- Appearance model

Layer Visibility

$\tau_{i,j}(x_i)$ foreground

$\pi_{i,j}(x_i)$ background

- See j-th foreground

$$P_j(x_i) = \tau_j(x_i)(1 - \sum_{l=1}^{j} \pi_l(x_i)) \cdot \prod_{s=1}^{j-1} [1 - \tau_s(x_i)]$$

In j-th foreground layer shape * not in a background layer * not in 1~(j-1)-th foreground layer

- See j-th background

$$P_{B,j}(x_i) = \pi_j(x_i) \cdot \prod_{k=1}^{j-1} (1 - \tau_k(x_i))$$

In j-th background layer shape * not in 1~(j-1)-th foreground layer

- Observe one background

$$P_B(x_i) = \sum_{j=1}^{L+1} \left[\pi_j(x_i) \cdot \prod_{k=1}^{j-1} (1 - \tau_k(x_i)) \right]$$

Sum all the possible background layers

Implementation of DLM

- Motion models
- Shape models
- Layer visibility
- Shape dynamics
- Appearance model

Shape Dynamics

- Assumption
  - Shapes don't change dramatically
  - Use constant velocity model
- Constant value Gaussian model

$$P(\tau_{t,j}(x_i) \mid \tau_{t-1,j}(x_i))$$
$$= \gamma + N(\tau_{t,j}(x_i) : \tau_{t-1,j}(\boxed{R(-\dot{\omega}_{t,j})(x_i - \dot{\mu}_{t,j}) / \dot{s}_{t,j}}), \sigma_\tau{}^2)$$

Shape map alignment

Implementation of DLM

- Motion models
- Shape models
- Layer visibility
- Shape dynamics
- Appearance model

Appearance Models

- $A_{t,j}(x_i)$
- Constant value over time
- Image model is a Gaussian distribution with A_t
- The temporal change of A_t is also a Guassian distribution

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

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

Summary: State estimation

$$P(\Lambda_t \mid \Lambda_{t-1}, I_t, \ldots, I_0)$$
$$= P_{order} \cdot P_{fg\_shape} \cdot P_{bg\_shape} \cdot P_{motion} \cdot P_{appearance} \cdot \prod_{i=1}^{N} (P_{bgo}(x_i) + P_{fgo}(x_i))$$

STEP 1: Find layer ordering

Go through all possible orderings and maximize the posterior probability

STEP 2: Motion estimation

Relaxing problem to $\prod_{i=1}^{n} (P_{bgo}(x_i) + P_{fgo}(x_i)) \cdot P_{motion}$

STEP 3: Foreground shape

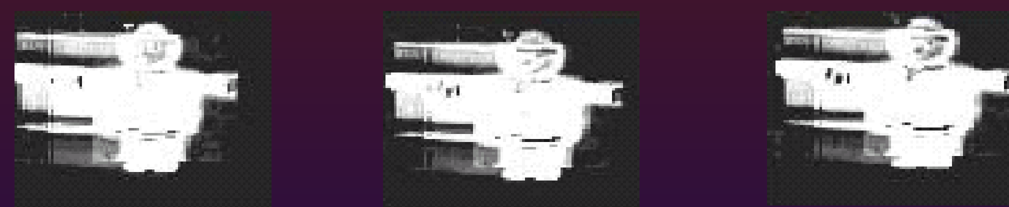
STEP 4: Background shape

STEP 5: Appearance estimation

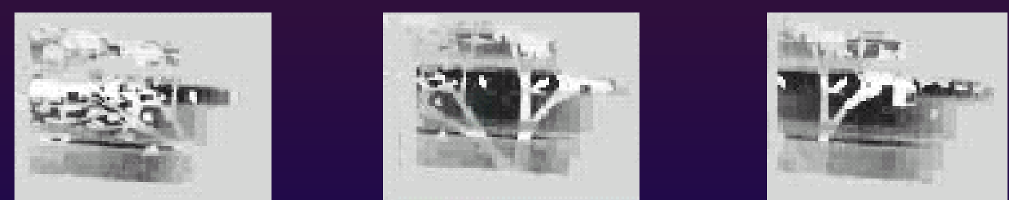
Search appearance value between current observation and the previous estimate

Results

Foreground shapes

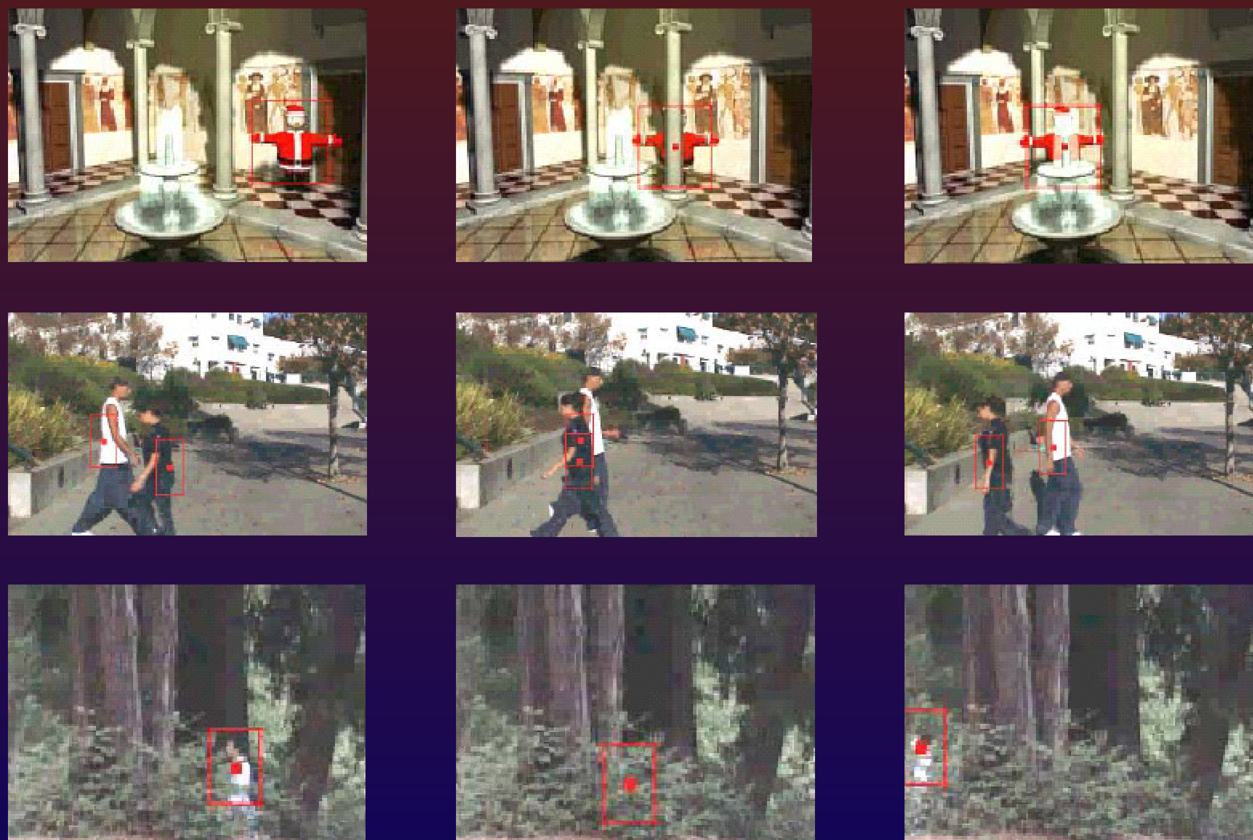


Background shapes



Foreground appearance

Results



Conclusion

Achievement

- Handling difficult occlusion problem
- Solving occlusion caused by the foreground and background objects

Future work

- Efficient optimization algorithms for optimal foreground layer ordering
- Flexible shape and motion model

Comparison

	Flexible Sprites	Background Layer Model
Moving Objects	Sprite layer	Foreground layer
Background	One layer	Multiple layer
Model parameters	Sprite mask Appearance	***Depth order, motion*** shape, appearance
Estimation	EM algorithm	EM for MAP + HMM

Discussion

- Why use background layers?

  Can we just take the background layers as foreground layers?

- When will this approach fail? Under which condition?
  - Observation noisy (Synthetic video)
  - Motion blur
- Does it work when there are multiple views?