Probabilistic Tracking in a Metric Space

Kentaro Toyama and Andrew Blake
Microsoft Research

Presentation prepared by:
Linus Luotsinen

Outline

- Introduction
- Modelling of images and observations
- Pattern theoretic tracking
- Learning
  - Learn mixture centers (exemplars)
  - Learn kernel parameters (observational likelihood)
  - Learn dynamic model (transition probabilities)
- Practical tracking
- Results
  - Human motion using curve based exemplars
  - Mouth using exemplars from raw image
- Conclusions

Introduction

- **Metric Mixture, M^2**
  - **Combine exemplars in metric space with probabilistic treatments**
  - **Models easily created directly from training set**
  - **Dynamic model to deal with occlusion**
- **Problems with other probabilistic approaches**
  - **Complex models**
  - **Training required for each object to be tracked**
  - **Difficult to fully automate**

University of Central Florida

Pattern Theoretic Tracking - Notation

Test images: $Z = \{z_1, z_2, ..., z_T\}$

Train images: $Z^* = \{z_1^*, z_2^*, ..., z_T^*\}$

Class is defined by a set of exemplars: $X = \{\tilde{x}_k, k = 1, 2, ..., K\}$

Geometrical transformation: $T_\alpha, \alpha \in A$ (known in advance)

Pattern theoretic tracking: $z \approx T_\alpha x$

State vector defined by: (α, k)

University of Central Florida

Metric Functions

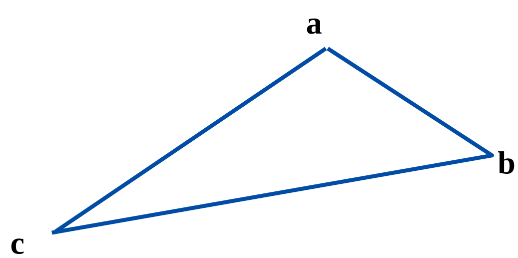
- **True metric function**
  - **All constraints**
- **Distance function**
  - **Without 3 and 4**

$$1)\ \rho(a,b) \geq 0, \forall\, a,b$$

$$2)\ \rho(a,b) = 0, iff\ \ a = b$$

$$\cancel{3)\ \rho(a,b) = \rho(b,a)}$$

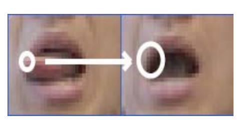
$$\cancel{4)\ \rho(a,b) + \rho(b,c) \geq \rho(a,c)}$$



University of Central Florida

Modelling of Images and Observations

- **Patches**
  - **Image sub-region**
  - **Shuffle distance function**
    - Distance with the most similar pixel in its neighborhood

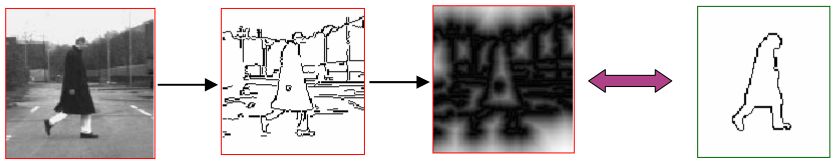


- **Curves**
  - **Edge maps**
  - **Chamfer distance function**
    - Distance to the nearest pixel in the binary images
    - See next slide!

University of Central Florida

Probabilistic Modelling of Images and Observations

- **Curves with Chamfer distance**



Original image — Feature image - I — Distance image – d_I — Exemplar - T

$$\rho_{\text{chamfer}}(I,T) = \frac{1}{|T|}\sum_{t\in T} d_I(t)$$

Not a true metric!

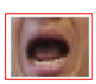
3) $\rho(I,T) \neq \rho(T,I)$

4) $\rho(A,B) + \rho(B,C) \ngeq \rho(A,C)$

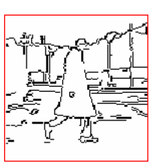
University of Central Florida

Pattern Theoretic Tracking

$$z \approx T_\alpha x$$



Observation

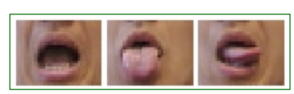


Geometrical Transform:

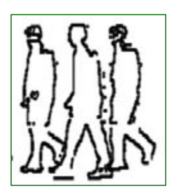
- Translation
- Affine
- Projective…

Exemplar (from a training set):

- Intensity images
- Feature images (edges, corners…)



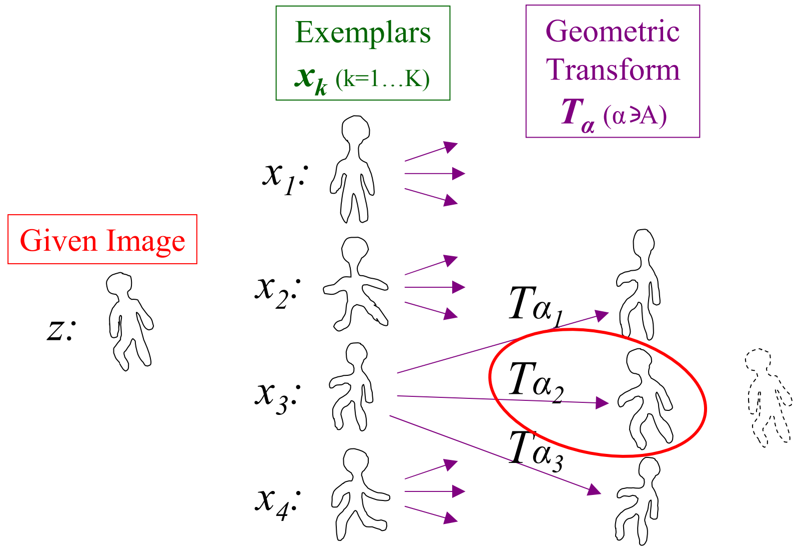
"patches"



"curves"

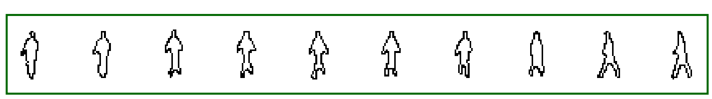
University of Central Florida

Pattern Theoretic Tracking

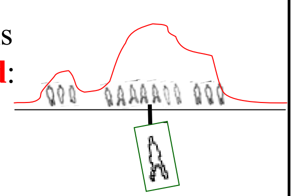


Pattern Theoretic Tracking - Learning

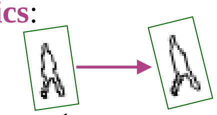
1) $\{\tilde{x}_k\}$ - A set of K **Exemplars**.



2) $\underset{X=(\alpha,k)}{p(z \mid X)}$ - Distribution of observations around $T_\alpha \tilde{x}_k$ – **Likelihood**:



3) $p(X_t \mid X_{t-1})$ - Prior about dependency between states – **Dynamics**:



University of Central Florida

Learning Mixture Centers

Goal - given M images (z_m), find K exemplars:

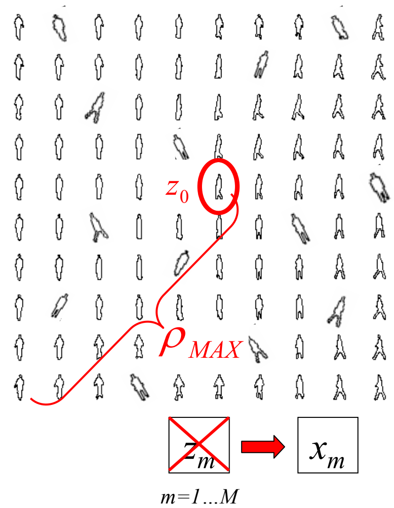
1) Find "central" exemplar -

$$z_0 = \arg\min_{z} \max_{z'} \rho(z, z')$$

2) "Align" other images to z_0 -

$$\alpha_m = \arg\min_{\alpha} \rho(T_\alpha^{-1} z_m, z_0)$$

$$x_m = T_\alpha^{-1} z_m$$



University of Central Florida

Learning Mixture Centers

3) Find K "distinct" exemplars -

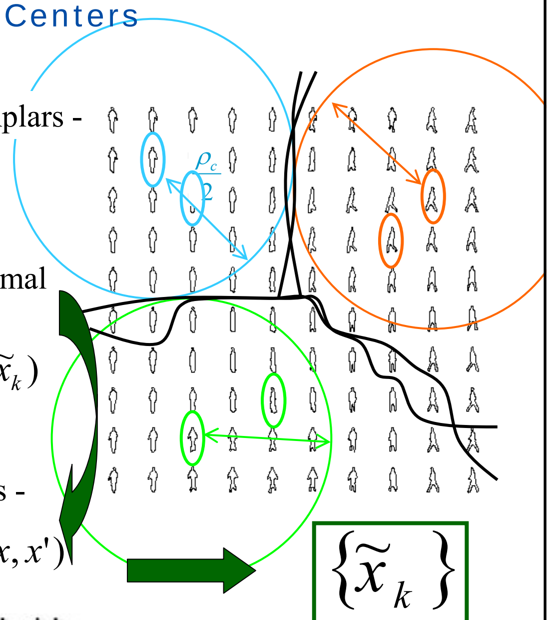
$$\rho(\tilde{x}_{k+1}, \tilde{x}_k) \approx \rho_c$$

4) Cluster the rest by minimal distance -

$$k_m = \arg\min_{k} \rho(x_m, \tilde{x}_k)$$

5) Find new representatives -

$$\tilde{x}_k = \arg\min_{x} \max_{x'} \rho(x, x')$$



University of Central Florida

Learning Kernel Parameters

1) Using a **"validation set"** find distances between images and their exemplars:

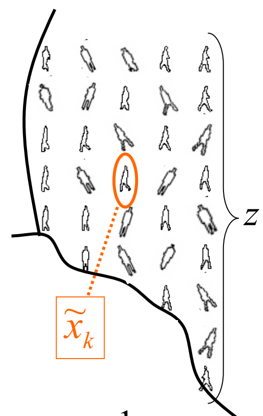
$$\rho(z, T_\alpha \tilde{x}_k)$$

2) Approx. distances as **chi-square**:

$$\rho(..) \sim \sigma^2 \chi_d^2 \quad \text{(to find } \boldsymbol{\sigma} \text{ and } \boldsymbol{d}\text{)}$$

3) Then the **observation likelihood** is:

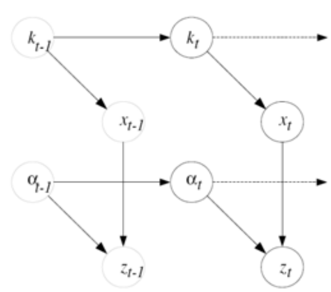
$$p(z \mid X) \propto \frac{1}{Z} \exp[-\lambda \rho(..)] \qquad Z \propto \sigma^d ; \lambda = \frac{1}{2\sigma^2}$$



University of Central Florida

Learning Dynamics

- **Learn a Markov matrix** M **for** $p(k_t | k_{t-1})$ **by histogramming transitions**
- **Run a first order auto-regressive process (ARP) for** $p(\alpha_t | \alpha_{t-1})$**, with coefficients calculated using the Yule-Walker algorithm**



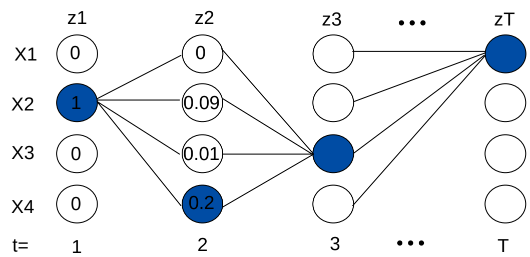
University of Central Florida

Practical Tracking

- **Forward algorithm**
  - $p_t(X_t) \equiv p_t(X_t \mid Z_1, Z_2, \ldots, Z_t)$

    $p_t(X_t) = \sum_{k_{t-1}} \int_{\alpha_{t-1}} p(z \mid X_t) p(X_t \mid X_{t-1}) p_{t-1}(X_{t-1})$

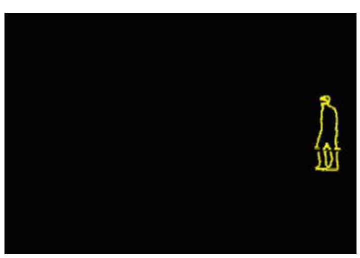
- **Results are chosen by**
  - $\hat{X}_t = \arg\max p_t(X_t)$



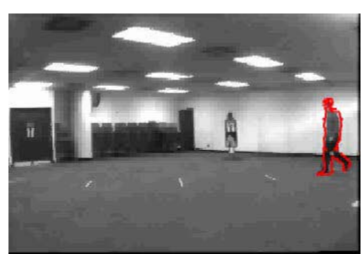
University of Central Florida

Results

- **Tracking human motion**
  - **Based on contour edges**
  - **Dynamics learned on 5 sequences of 100 frames each**



Exemplars

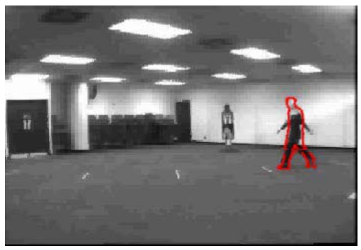


Same person, motion not seen in training sequence

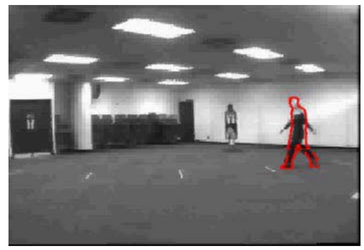
University of Central Florida

Results

- Tracking human motion
  - Based on contour edges
  - Dynamics learned on 5 sequences of 100 frames each



Different person

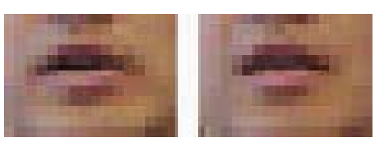


Different person with occlusion (power of dynamic model)

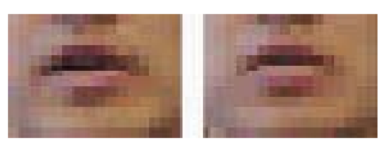
University of Central Florida

Results

- Tracking person’s mouth motion
  - Based on raw pixel values
  - Training sequence was 210 frames captured at 30Hz
  - Exemplar set was 30 (K=30)
- Left image show test sequence
- Right image show maximum a posteriori



Using L2 distance

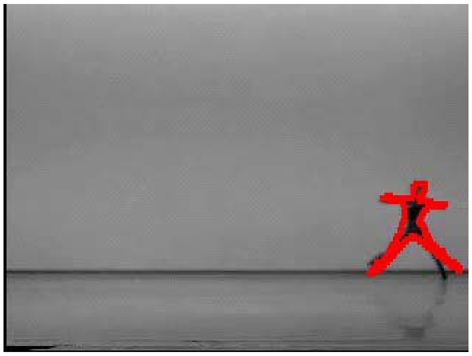


Using shuffle distance

University of Central Florida

Results

- Tracking ballerina
  - Larger exemplar sets (K=300)



Conclusions

- Metric Mixture (M^2) Model
  - Easier to fully automate learning
  - Avoid explicit parametric models to describe target objects
- Generality
  - Metrics can be chosen without significant restrictions
- Temporal fusion of information for occlusion recovery
  - Bayesian networks

References

- [1] Kentaro Toyama, Andrew Blake, Probabilistic Tracking with Exemplars in a Metric Space, International Journal of Computer Vision, Volume 48, Issue 1, Marr Prize Special Issue, Pages: 9–19, 2002, ISSN:0920-5691.
- [2] Jongwoo Lim, CSE 252C: Selected Topics in Vision & Learning. http://www-cse.ucsd.edu/classes/fa02/cse252c/
- [3] Eli Schechtman and Neer Saad, Advanced topics in computer and human vision. http://www.wisdom.weizmann.ac.il/~armin/AdvVision02/course.html